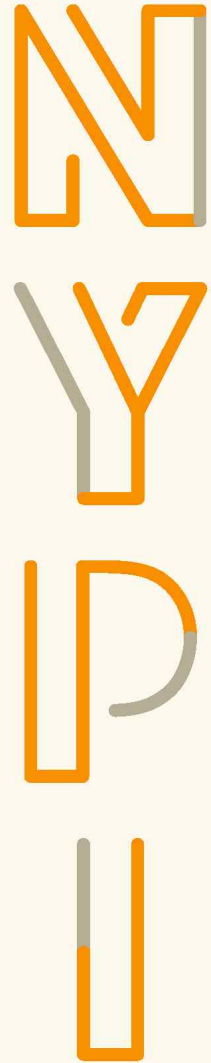


연구보고 20-R08

청소년 빅데이터 체계 구축 및 활용방안 연구

서정아 성윤숙 송태민



연구보고 20-R08

청소년 빅데이터 체계 구축 및 활용방안 연구

저 자 서정아, 성윤숙, 송태민

연 구 진 연구책임자_ 서정아(한국청소년정책연구원 선임연구위원)
공동연구원_ 성윤숙(한국청소년정책연구원 선임연구위원)
_ 송태민(삼육대학교 교수)
연구보조원_ 이유정(한국청소년정책연구원 연구사업운영원)

급속한 사회환경 변화와 청소년 현안의 복잡성 증가로 인해 다양한 정형, 비정형 빅데이터에 기반한 청소년 정책수립과 미래예견적 의사결정의 중요성이 강조되고 있습니다.

본 연구는 이러한 중요성에 기반하여 문헌고찰, 분석방법에 관한 콜로키움, 외국의 빅데이터 체계 구축 및 활용 사례 검토, 청소년 핵심 영역 온톨로지 개발, 전문가 심층면접, 빅데이터 체계 구축 및 활용 방안 등에 관한 전문가 조사, 비정형 빅데이터 분석, 정형 빅데이터 분석 등을 통해 청소년 대상 빅데이터 체계 구축의 가능성을 파악하고 방안을 제시합니다.

청소년 연령 규정의 다양성과 모호성, 청소년 관련 이슈에 대한 위기 중심·잔여적 접근의 한계, 비정형 빅데이터의 광범위성 등을 고려하여 연구의 범위는 광의의 보편적 청소년복지 관점에서 청소년기본법 상의 청소년 연령인 9세~24세 청소년을 대상으로 하는 공공 정형 빅데이터, 소셜 비정형 빅데이터에 초점을 맞춥니다.

본 연구가 상대적으로 빅데이터에 대한 논의가 활발하게 진행되지 못한 청소년 분야에서 빅데이터에 대한 본격적인 논의가 이루어지는 중요한 계기가 될 수 있기를 희망합니다.

2020년 12월

한국청소년정책연구원

원장 김 현 철

국문초록

본 연구에서는 청소년 대상 빅데이터 체계 구축의 가능성을 파악하고 이를 바탕으로 체계 구축을 위한 제안점을 도출함으로써 청소년 빅데이터 체계 구축을 위한 기초자료를 제시함은 물론 정형 빅데이터와 비정형 소셜 빅데이터를 활용한 분석을 시도함으로써 청소년관련 이슈에 대해 어떻게 예측이 가능한지를 확인해 보고 청소년빅데이터 체계가 추후 청소년관련 미래변화나 정책수요를 예측하고 위험요소를 조기 발견, 예방하는데 어떻게 활용될 수 있는지 등에 관한 구체적 방안을 도출, 제시하고자 한다.

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 활용하였다.

주요 연구결과는 다음과 같다.

첫째, 전문가 심층면접 결과 청소년 빅데이터가 실질적으로 다루어야 할 사안과 관련하여 응답자들은 크게 생애설계적 관점과 생애수정적 관점을 강조하였다. 둘째, 전문가 설문조사 결과 공공 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항으로 개인정보 보호로 인한 접근의 제한성이 37.5%로 가장 많았다.

연구결과를 바탕으로 청소년 빅데이터 체계 구축 방안 및 활용방안을 제시하였다.

연구요약

1. 연구목적

- 본 연구에서는 청소년 대상 빅데이터 체계 구축의 가능성을 파악하고 이를 바탕으로 체계 구축을 위한 제안점을 도출함으로써 청소년 빅데이터 체계 구축을 위한 기초자료를 제시함은 물론, 정형 빅데이터와 비정형 소셜 빅데이터를 활용한 분석을 시도함으로써 청소년관련 이슈에 대해 어떻게 예측이 가능한지를 확인해 보고 청소년빅데이터 체계가 추후 청소년관련 미래변화나 정책수요를 예측하고 위험요소를 조기 발견, 예방하는데 어떻게 활용될 수 있는지 등에 관한 구체적 방안을 도출, 제시하고자 한다.

2. 연구방법

- 연구방법으로 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 심층면접, 전문가 설문조사, 정책협의회 등을 활용하였다.

3. 주요결과

1) 이론적 배경

- 다양한 변화와 그에 따른 제반 문제, 그리고 앞에서 살펴본 이론적 배경을 종합적으로 고려할 때 청소년 삶의 변화에 대응하는 시의적절한 대책설계를 위해 청소년 빅데이터 구축은 필수적인 것으로 보인다. 무엇보다 국가 차원의

빅데이터 전략계획이 필요하고 학계, 기업, 시민단체, 일반국민 등 다양한 이해관계자의 참여와 협력을 가능하게 할 거버넌스, 그리고 청소년 빅데이터 연구를 리드하는 전담기관으로서의 컨트롤 타워가 필요하다. 이를 토대로 개인정보보호를 위한 법, 제도 정비가 이루어지고 동시에 데이터의 체계적 수집과 지속적 자료 축적 그리고 데이터 리터러시 강화를 위한 다양한 노력이 실질적 관심과 의지로 이루어진다면 청소년 빅데이터 구축과 실질적 활용가능성이 담보될 수 있을 것이다.

2) 청소년 빅데이터 분류 체계

- 빅데이터 체계를 구축하기 위해서는 무엇보다 청소년 빅데이터의 전체 내용이 어떠한 지형을 이루고 있는지 파악할 필요가 있다. 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 있으며 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 있다. 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 있다. 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 있다. 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 수정보완 방안을 제시하였다.

3) 외국의 아동청소년 빅데이터 체계 구축 사례

- 첫째, 빅데이터 일반 뿐만 아니라 청소년 빅데이터 체계구축을 위한 정부주도의 노력이 필요하다. 먼저 호주는 아동 및 청소년을 위한 빅데이터 정책, 전략계획, 연구 등을 위하여 국가 차원에서 공공서비스를 위한 빅데이터 전략계획을 수립하고 추진하는 것으로 확인하였다. 영국의 경우도 일찍부터 데이터의 가치에 주목하며 정부 주도의 데이터 정책을 추진해 왔다. 2000년 정보공개법 제정을 시작으로 2009년 「스마트 정부 구현을 위한 실행계획서」, 2012년 「오픈 데이터 백서」, 2014년 「오픈 데이터 전략 2014-2016」 등을 발표하여 공공 데이터 개방 정책을 펼치며 정부 운영의 효율성과 투명성, 궁극적으로 책무성을 제고하고자 하였다. 이후에는 데이터 분석과 활용에 집중하며 2013년 범국가적 「데이터 역량 강화 전략」, 2017년 「디지털 전략」 등을 추진하였고 현재는 「국가 데이터 전략」을 고안하는 중이다. 이처럼 영국 정부는 (빅)데이터를 미래 영국을 이끌어갈 주요한 성장동력으로 여기며, 관련 법제를 마련하고 정책을 추진하는데 매우 적극적인 모습이다.
- 둘째, 빅데이터 구축과 활용을 위한 거버넌스가 요구된다. 영국의 경우 처럼 학계, 기업, 시민단체, 일반 국민 등 다양한 이해관계자의 참여와 협력이 필요하다. 예를 들어, 「국가 데이터 전략」 수립을 주관하는 영국

의 DCMS는 정책의 구체적 설계에 앞서 공공 및 민간(기업, 시민단체, 학교와 연구기관 등) 부문 그리고 일반 국민을 대상으로 현황조사(Call for Evidence)를 실시하여 사람, 경제, 정부의 세 차원에서 데이터 관련 인식과 관심, 실태, 욕구 및 현안 등을 파악하였다. 또한 DCMS는 공공부문과 기업, 시민단체 등 250여 곳에서 참가자들을 모아 20회 이상의 원탁회의도 진행하며 다양한 이해관계자의 참여를 독려하고 그 의견을 생생하게 수렴하고 있다.

4) 전문가 심층면접

- 응답자들은 빅데이터가 기존 데이터베이스의 처리방식을 초월하는 방대한 양의 정형 및 비정형 데이터라는 일반적 정의에 모두 공감하고 있었다. 또한 중요성과 관련해서는 모든 일상의 데이터화, 데이터의 확장과 연계 그리고 데이터 처리 기술의 발전 등으로 사회현상, 특히 청소년이나 청소년 활동과 관련된 사안의 보다 명시적인 확인, 정밀한 파악과 분석이 가능하게 되었고, 이를 통해 실증적 데이터를 기반한 이론이나 논리의 발견과 예측이 가능해졌음을 지적하였다.
- 청소년 빅데이터가 실질적으로 다루어야 할 사안에 대하여 응답자들은 크게 생애설계적 관점과 생애수정적 관점의 두 관점에서 접근하는 모습을 보여주었다. 생애설계적 관점, 즉 청소년의 인생을 생산적이고 긍정적으로 구성하는 측면에 중심을 둔 입장은 증분류로 진로, 주관적 복지감, 심리, 안전, 인구, 가족, 건강, 교육, 여가, 문화, 시간 이용, 관계 및 사회참여 등등을 언급하였다.
- 청소년 빅데이터 체계 구축 및 활용의 장점에 대해 응답자의 다수는 데이터 활용시의 효율성 제고를 언급하였다. 특히 데이터 전처리에 필요한 비용과

시간절감 그리고 각 기관마다 보유하고 있는 자료의 특성들을 고려한 데이터 연계가능성의 장점을 지적하며 이를 통해 청소년 문제에 대한 선제적 대응, 예방과 청소년의 잠재력 개발 등 다양한 측면에서의 정책적 활용이 가능하리라고 보았다.

5) 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 응답하였다. 전문가별로는 현장전문가, 공무원은 66.7%, 교수, 연구원은 46.4%가 사용 경험이 있다고 응답하였다.
- 공공 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항을 살펴보면 전체적으로 개인정보보호로 인한 접근의 제한성이 37.5%로 가장 많았으며 다음으로 데이터 형식간의 통일성 부재로 인한 분석상의 어려움, 데이터 품질 및 개선사항에 대한 의견 수렴 부족, 홍보 부족으로 인한 데이터 관련 정보부재, 데이터 세부 사항, 범례 등에 대한 설명 부족, 원스톱 서비스 또는 센터의 부재(한 곳에 모아놓기를 희망), 원데이터를 분석 가능한 데이터로 변환하는 과정 불편이 각 12.5%, 데이터 항목의 누락이나 중복 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%였다.
- 민간 영역 빅데이터를 사용한 경험이 있는 경우 사용한 민간 영역 빅데이터의 유형을 살펴보면 전체적으로 국내일간지/온라인뉴스채널, SK텔레콤+통계

청 유동생활 인구분석, 복합데이터(뉴스+게시판+자료실+SNS), 다음/네이버카페가 27.3%로 가장 많았으며 다음으로 금융빅데이터플랫폼(BC 카드) 18.2%, 네이버지식인, 부동산114, 네이버랩, 트위터, 유튜브, 다음SNS, 빅카인즈, 구글트렌드, 인스타그램, 구글크롬로그파일(log file), 네이버블로그, 페이스북, 각종온라인게시판/댓글, 네이트판이 각 9.1%로 나타났다.

- 민간 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항을 살펴보면 전체적으로 민간데이터 구매방법이 어렵거나 가격이 높음이 33.3%로 가장 많았으며 다음으로 공개데이터 양, 기간의 한계, 수집 및 분석에 큰 어려움이없었음, 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 세부사항 특정하기 어려움, 분석의 한계(예: 트렌드 분석은 가능하나 궤적 연구는 어려운 것 등)가 각 8.3%였다. 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 세부사항 특정하기 어려움이 각 25.0%였다. 교수, 연구원의 경우 수집 및 분석에 큰 어려움이 없었음, 공공데이터에 비해 데이터의 신뢰성, 타당성 확인 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 놀라움, 걱정, 분석의 한계(예: 트렌드분석은 가능하나 궤적연구는 어려운 것 등)가 각 12.5%였다.

4. 정책제언

- 청소년 빅데이터 체계 구축을 위한 비전 및 전략 도출
- 청소년 빅데이터 체계 구축 및 활용을 위한 세부 정책 과제 제시
- 미래예측을 위해 우선적으로 구축되어야 할 청소년 관련 빅데이터 세부 정책 과제 제안
- 빅데이터 온톨로지 체계 세부내용 개선 방안 제안

청소년 빅데이터 체계 구축 및 활용방안 연구

연구보고 20-R08

I. 서론

1. 연구의 필요성 및 목적	3
2. 주요 연구내용	6
1) 빅데이터에 관한 이론적 배경 검토	6
2) 청소년 지표체계와 빅데이터 온톨로지 개발	6
3) 외국의 아동·청소년 빅데이터 구축 사례 검토	7
4) 빅데이터 체계 구축 및 활용방안 모색을 위한 전문가 조사(심층/설문)	7
5) 정형 빅데이터 사례 분석 실시	8
6) 비정형 빅데이터 사례 분석 실시	8
7) 청소년 빅데이터 체계 구축 및 활용을 위한 비전 및 전략, 세부 정책과제 도출	8
3. 연구방법	9
1) 문헌연구	9
2) 전문가 자문	9
3) 양적 연구	10

4) 질적 연구	10
5) 기타 연구방법	11

II. 이론적 배경

1. 빅데이터의 개요	15
1) 청소년 빅데이터의 개념	15
2) 청소년 빅데이터의 필요성	16
3) 국내 아동·청소년 빅데이터 구축 및 활용 현황	18
2. 아동·청소년 빅데이터 관련 선행연구	28
3. 빅데이터의 활용 및 분석	32
1) 비정형 빅데이터의 활용	32
2) 비정형 빅데이터의 분석	34
4. 빅데이터 개인정보보호 방안	35
1) 개인정보, 프라이버시, 익명성의 개요	36
2) 빅데이터 활용에서의 개인정보보호방안	39
3) 시사점	42

III. 청소년 빅데이터 분류 체계

1. 청소년 빅데이터 분류 체계 기준 자료 검토	45
2. 청소년 빅데이터 분류 체계 관련 개선 의견	58
3. 통계청 아동·청소년 관련 통계 현황 조사	66

IV. 외국의 아동·청소년 빅데이터 체계 구축 사례

1. 미국의 아동·청소년 빅데이터 체계 구축 및 활용 사례 ..	69
1) 서론	69
2) 미국의 빅데이터 체계 현황	71
3) 정책 사례	82
4) 우리나라의 정책적 시사점	96
2. 영국의 아동·청소년 빅데이터 체계 구축 및 활용 사례 ..	103
1) 데이터 구축 및 공개에 관한 주요 정책	103
2) 빅데이터 관련 법제	117
3) 빅데이터 정책 추진 체계	120

4) 빅데이터 기반 정책 연구 체계: Administrative Data Research UK (ADR UK)	125
5) 빅데이터 기반 아동·청소년 연구 사례: ADR UK의 아동·청소년 연구 프로젝트	129
6) 시사점	135
3. 호주의 아동·청소년 빅데이터 체계 구축 및 활용 사례 ..	139
1) 호주의 빅데이터 정책	139
2) 호주의 청소년 빅데이터 현황	165
3) 호주의 청소년 빅데이터 적용 사례	173
4) 우리나라에 대한 시사점	177
4. 외국의 아동·청소년 빅데이터 체계 구축·활용 사례 소결 및 시사점	179

V. 전문가 심층면접

1. 심층면접 개요	189
1) 조사 목적	189
2) 연구 대상	189
3) 조사도구	190
4) 연구절차	192
2. 심층면접에 대한 분석	193
1) 빅데이터 개념 및 중요성	193
2) 청소년 빅데이터의 개념	200
3) 국내외 빅데이터 체계 구축 상황과 우수사례	204
4) 빅데이터 활용 경험	208
5) 활용 가능한 정형, 비정형 빅데이터 유형	212
6) 청소년 대상 빅데이터 지표 체계 분류 영역	216
7) 청소년 대상 빅데이터 활용 방안	228
8) 청소년 대상 빅데이터 체계 구축 방안	251
9) 청소년 빅데이터 체계 구축 및 활용의 발전방향 ..	265
3. 소결	269

VI. 전문가 조사

1. 조사대상 일반특성	279
--------------------	-----

2. 빅데이터 사용 경험	280
1) 공공 영역 빅데이터 사용 연구·분석 경험	280
2) 민간 영역 빅데이터 사용 연구·분석 경험	287
3. 빅데이터 관련 욕구	292
1) 빅데이터에 대한 관심 및 활용 희망 여부	292
2) 빅데이터 활용 희망하지 않는 경우 이유	293

Ⅶ. 결론 및 제언

1. 청소년 빅데이터 체계 구축을 위한 비전 및 전략 ...	297
2. 청소년 빅데이터 체계 구축 및 활용을 위한 세부 정책과제 도출	299
1) 빅데이터 활용 향상을 위한 세부 정책과제	299
2) 미래예측을 위해 우선적으로 구축되어야 할 청소년 관련 빅데이터 세부 영역	300
3) 아동·청소년 분야 빅데이터 체계 구축을 위한 세부 정책과제	301
3. 청소년 빅데이터 온톨로지(안) 제시	304

참고문헌	307
-------------------	------------

부 록	321
------------------	------------

Abstract	347
-----------------------	------------

표 목차

표 II-1 빅데이터 분석기반의 위기청소년 예측 및 적시대응 기술개발 연계 사례	41
표 III-1 UNICEF 아동 웰빙 지표 체계	46
표 III-2 아동·청소년 핵심지표 영역체계(안)	48
표 IV-1 미국 Data.gov의 Youth Program 검색에 따른 Datasets 종류(2020년 8월)	74
표 IV-2 미국 2020년 KIDS COUNT에 나타난 미국아동·청소년 추세(2020년 8월)	80
표 IV-3 영국의 데이터 구축 및 공개 관련 주요 정책	103
표 IV-4 기업혁신기술부(BIS)의 공개 데이터(예시)	106
표 IV-5 데이터 이용 개선을 위한 중점 과제(2017 - 2020)	112
표 IV-6 국가 데이터 전략 수립을 위한 현황조사 질문	113
표 IV-7 ADR UK의 연구주제 및 프로젝트	128
표 IV-8 ADR UK가 승인, 지원하고 있는 아동·청소년 관련 연구 프로젝트	129
표 IV-9 변화를 위한 청사진 (1/2)	159
표 IV-10 변화를 위한 청사진 (2/2)	160
표 IV-11 전략적 조정(연계)의 주요 측면들	162
표 IV-12 데이터의 기능 및 역량	164
표 IV-13 호주의 청소년 개방 데이터 set의 수 및 유형 (data.gov.au) ..	166
표 IV-14 개방데이터 자체 평가표	167
표 V-1 심층면접 참여자의 인적 특성	190
표 V-2 심층면접 질문지 구성 내용	191
표 VI-1 전문가 조사대상 일반특성	279
표 VI-2 공공 영역 데이터 사용 경험 유무	280
표 VI-3 사용한 공공 영역 빅데이터(기관 기준)(다중응답)	281
표 VI-4 공공 영역 빅데이터로 분석한 내용/주제(다중응답)	283

표 VI-5 공공 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항 (다중응답)	285
표 VI-6 민간 영역 빅데이터 사용 경험 유무	287
표 VI-7 사용한 민간 영역 빅데이터 유형(다중응답)	288
표 VI-8 민간 영역 빅데이터로 분석한 내용/주제 영역(다중응답)	289
표 VI-9 민간 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항 (다중응답)	290
표 VI-10 빅데이터에 대한 관심	292
표 VI-11 1년 이내 빅데이터 활용 희망 여부	293
표 VI-12 빅데이터 활용 희망하지 않는 경우 이유	294

그림 목차

그림 II-1	청소년 빅데이터 구축의 필요성	17
그림 II-2	공공데이터포털 홈페이지(https://www.data.go.kr)에서 제공하는 청소년 통계 화면	18
그림 II-3	과학기술정보통신부 한국정보화진흥원 빅데이터센터(K-ICT) 홈페이지(https://kbig.kr/portal/kbig)	20
그림 II-4	행정안전부 우리의 미래, 빅데이터 홈페이지 (http://www.bigdata.go.kr)	21
그림 II-5	행정안전부 데이터 분석 결과	22
그림 II-6	보건복지부 국민건강보험공단 홈페이지 (https://www.nhis.or.kr/retrieveHomeMain.xx)	23
그림 II-7	부산광역시 빅데이터포털 홈페이지 (https://bigdata.busan.go.kr/index.busan)	24
그림 II-8	한국청소년정책연구원 홈페이지 (https://www.nypi.re.kr/archive)	25
그림 II-9	보건사회연구원 홈페이지(https://www.kihasa.re.kr/)	25
그림 II-10	한국행정연구원 홈페이지 (https://www.kipa.re.kr/site/kipa/main.do)	26
그림 II-11	개인에 관한 정보와 개인정보 및 프라이버시의 관계	37
그림 II-12	개인정보와 프라이버시에 해당하는 정보와의 관계	38
그림 IV-1	영국 오픈 데이터 포털 'data.gov.uk'	108
그림 IV-2	국가 데이터 전략 수립을 위한 원탁회의	117
그림 IV-3	Policy Lab의 정책 디자인 프레임워크	123
그림 IV-4	ODI의 변화 이론(Theory of Change)	125
그림 IV-5	ADR UK의 파트너십	127
그림 VI-1	청소년 빅데이터 체계 구축을 위한 비전 및 전략	297

○ ————— 제1장 서론

- 1. 연구의 필요성 및 목적
- 2. 주요 연구내용
- 3. 연구방법

1. 연구의 필요성 및 목적¹⁾

4차 산업혁명 시대에 국가 경쟁력의 원천이자 급격한 사회변화에 대응하고 주요 현안을 해결하기 위한 방안으로 빅데이터의 활용가치가 더욱 높아지고 있다. '데이터 경제(Data Economy)'의 확산으로 가늠할 수 없을 정도로 방대하고 다양한 데이터가 빠르게 생산되고 있으며 이러한 빅데이터의 생산은 앞으로도 폭발적으로 증가할 것으로 예측된다. OECD와 EU 등 국제기구(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 있다(이슈 & NDSL, 미국 연방정부, 빅데이터 연구개발 전략계획 공개, 2020.2.10. 검색). 증거기반 정책수립과 미래 예견적 국정관리 측면에서도 빅데이터 구축 및 활용의 필요성이 강조되고 있다. 최근 정부는 데이터 산업 활성화 전략을 마련하고 2019년부터 빅데이터 플랫폼 및 센터 구축 사업을 추진 중에 있으며(대통령직속4차산업혁명위원회, 데이터 산업 활성화 전략 데이터 산업 활성화 전략, 2020.2.10. 검색) 통계청과 행정안전부를 중심으로 공공데이터를 확보하고 개방해 나가는 추세에 있다. 특히 청소년은 인터넷이나 모바일에 기반한 디지털 기기 사용이 일상화 되어 있고, 디지털 생태계에서 살아가는 청소년의 삶의 양식이 이전 세대와 다른 양상을 나타내는 만큼

1) '제1장 서론'은 한국청소년정책연구원 서정아 선임연구위원이 작성하였다.

청소년에 대한 정확하고 적시성 있는 정책 마련을 위해 특히 소셜 빅데이터를 활용한 청소년 연구가 필수적이다. 사회 환경의 급속한 변화와 청소년 현안의 복잡성 증가에 따라 다양한 정형, 비정형 빅데이터에 기반한 청소년 정책수립과 미래예견적 의사결정의 중요성이 강조되고 있는 것²⁾이다.

빅데이터의 중요성에 대한 전지구적 관심으로 다양한 빅데이터 관련 국·내외 연구(강희정, 2015; 송태민, 2012; 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, 배충효 역, 2015; 후나오 노부오 저, 김성재 역, 2014) 보건의료 등 일부 영역에서 빅데이터 활용을 위한 체계 마련 노력이 이루어지고 있다(강희정, 2015). 그러나 청소년 분야에서 정형, 비정형 빅데이터 체계를 파악하고 이를 구축, 활용하고자 하는 노력은 아직까지 이루어지지 못하고 있다. 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 있다.

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

2) 최근 들어 미래 환경 변화 감지를 위한 다양한 시도가 이루어지고 있는데 그 중 가장 많은 주목을 받고 있는 것이 '소셜 빅데이터 기반 미래신호 예측방법론'을 통해 미래의 변화를 예측할 수 있는 약신호(weak signal)를 탐지하는 것이다. 약신호는 시간이 흐름에 따라 강신호(strong signal)로 다시 트렌드나 메가 트렌드로 변화한다(박찬국·김현재, 2015).

바탕으로 체계 구축을 위한 제안점을 도출함으로써 청소년 빅데이터 체계 구축을 위한 기초자료를 제시함은 물론 정형 빅데이터와 비정형 소셜 빅데이터를 활용한 분석을 시도함으로써 중요한 청소년관련 이슈에 대해 어떻게 예측이 가능한지를 확인해 보고 청소년빅데이터 체계가 추후 청소년관련 미래변화나 정책수요를 예측하고 위험요소를 조기 발견, 예방하는데 어떻게 활용될 수 있는 지 등에 관한 구체적 방안을 도출, 제시하고자 한다. 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 (㉠) 청소년기본법상의 청소년 연령인 9세~24세 청소년을 주된 대상으로 하는 (㉡) 공공영역 정형 빅데이터와 (㉢) 소셜 비정형 빅데이터 분석에 초점을 맞추려 한다. 이는 빅데이터에 대한 선행연구가 모든 국민을 대상으로 하는 공공 서비스 부문 정형 빅데이터의 중요성을 강조하고 이를 우선 검토한 것(이현웅·김종업·최현재, 2018), 다양한 비정형 빅데이터 중에서도 특히 소셜 빅데이터의 중요성에 주목하고 이에 초점을 맞춘 것(송태민·송주영, 2016)과 일맥상통한다. 빅데이터의 형태는 영상, 이미지, 문자 등 다양하게 나타나는데 본 연구에서는 특히 문자로 나타나는 빅데이터에 초점을 맞춘다. 또한 본 연구에서는 기존의 청소년 관련 빅데이터 연구가 주로 위기, 비행 등 부정적 측면과 잔여적 복지에 초점을 맞추고 있어(고수정, 2018; 안명숙, 2018) 모든 청소년에게 적용하는데 한계가 있으며 청소년을 대상으로 한 서비스가 보다 보편적 시각에서 접근되어야 한다는 최근의 청소년복지관련 흐름에 발맞추어 (㉣) 보다 광범위하고 긍정적인 광의의(보편적) 청소년복지 개념에서 청소년관련 핵심개념을 설정하고 이에 대한 온톨로지 분석과 논의를 해 나가자 한다.

2. 주요 연구내용

1) 빅데이터에 관한 이론적 배경 검토

국내·외 문헌, 인터넷 자료를 활용하여 빅데이터에 대한 이론적 배경을 검토한다. 세부 내용은 다음과 같다.

- 빅데이터 개요: 정의, 필요성, 유형, 현황(자료 위치, 구조) 등
- 아동·청소년 관련 빅데이터 관련 국내·외 연구 동향
- 빅데이터 분석 과정: 빅데이터 저장, 정제, 처리, 분석, 표현과정, 유의사항 등
- 빅데이터 관련 법제도적 이슈 분석: 개인정보보호 관련 이슈, 지적재산권 관련 이슈, 빅데이터 알고리즘 규제 이슈 분석 등

2) 청소년 지표체계와 빅데이터 온톨로지 개발

빅데이터 체계를 구축하기 위해서는 무엇보다 청소년 빅데이터의 전체 내용이 어떠한 지형을 이루고 있는지 파악할 필요가 있다. 그러나 이 작업을 하기 이전에 선행되어야 할 사항은 청소년 관련 공공영역 정형데이터, 소셜 비정형데이터를 검색하여 분류할 지표체계를 만드는 것이다. 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 있으며 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 있다. 국제적 추세와 한국적 상황을 모두 고려하기 위해 국제기구의 지표체계와 한국의 지표체계를 조합하여 한국형 지표

체계를 확정하는 과정이 선행되어야 할 것으로 보인다. 이렇게 확정된 지표체계를 바탕으로 각 지표영역별로 어떠한 활용 가능한 빅데이터가 있는지 파악한다. 온톨로지는 청소년 빅데이터의 유형, 조건, 세부 개념 등을 도식화(map)하기 위한 빅데이터 분류 체계이다. 온톨로지 개발을 위해 공공영역의 정형 빅데이터를 중심으로 데이터를 검토함으로써 빅데이터 체계의 가능성을 타진하고 추가 데이터 마련이 필요한 영역, 데이터 간의 융합이 이루어질 수 있는 가능성 등을 파악할 수 있을 것이다.

3) 외국의 아동·청소년 빅데이터 구축 사례 검토

외국이 아동·청소년 빅데이터 구축을 위해 어떠한 노력을 기울이고 있는지 정책 추진 동향과 빅데이터 구축 및 활용 사례를 분석한다. 대상 국가는 미국, 영국, 호주이며 이는 1차 전문가 자문을 토대로 선정된 것이다.

4) 빅데이터 체계 구축 및 활용방안 모색을 위한 전문가 조사(심층/설문)

전문가 심층면접과 자문, 설문조사 등을 통해 청소년 빅데이터 체계 구축 및 활용방안, 체계 구축을 위한 우선사항 등을 파악한다. 이 과정에 학계, 공공부문 전문가를 포함하였다. 파악한 내용은 다음과 같다.

- 개요: 빅데이터의 중요성, 빅데이터란?
- 현황: 활용가능한 빅데이터 유형, 빅데이터를 통해 도출해야 하는 사항, 우선적으로 분석해야 할 인과관계
- 미래전망: 빅데이터 체계 구축의 구체적 방법, 빅데이터의 활용방법, 체계 구축의 우선순위, 중요도에 대한 인식 등

5) 정형 빅데이터 사례 분석 실시

공공영역 정형 빅데이터를 어떻게 분석할 수 있는지 수집 및 정제, 데이터 분석 방법, 분석 결과, 해석 방법 등을 알아보기 위해 공공영역 정형 빅데이터 사례를 분석해보고 그 결과를 제시하였다. 분석주제는 청소년 행복 결정 요인이며 데이터 분석 보고서 2로 별도 제시하였다.

6) 비정형 빅데이터 사례 분석 실시

비정형 빅데이터가 실제로 어떻게 분석되어 지는지 수집 및 정제, 데이터 분석 방법, 분석 결과, 해석 방법 등을 파악하기 위해 소셜 비정형 데이터를 통해 추출, 수집된 2차 데이터를 활용하여 비정형 빅데이터 사례를 분석해보고 그 결과를 제시하였다. 주제는 청소년 건강 중에서 비만에 관한 것이며 데이터 분석 보고서 1로 별도 제시하였다.

7) 청소년 빅데이터 체계 구축 및 활용을 위한 비전 및 전략, 세부 정책과제 도출

연구결과를 통해 도출된 내용을 바탕으로 한국의 빅데이터 체계 구축 상황을 객관적으로 파악하고 이를 바탕으로 청소년 빅데이터 구축 및 활용을 위한 구체적인 비전 및 전략, 세부 정책과제를 도출하였다.

3. 연구방법

1) 문헌연구

국내·외 선행연구와 인터넷 사이트 등을 검토함으로써 청소년 빅데이터 현황 및 특성을 분석하고 외국의 빅데이터 정책 추진 동향 및 활용 사례를 파악한다. 청소년 빅데이터에 대한 정책, 사업, 법제도, 기술 동향 또한 검토한다.

2) 전문가 자문

연구과정 전반을 거쳐 전문가 자문을 실시하였다. 세부사항은 다음과 같다.

횟수	자문영역	자문진	비고
2회	<ul style="list-style-type: none"> ○ 실행계획서 작성 및 연구진행 전반에 관한 자문(1차) <ul style="list-style-type: none"> - 2020.2.3(월) 실시(장소: 서울역 회의실) - 자문내용: 연구 진행 절차와 연구방법 전반 - 자문진: 빅데이터 전문가 2명, 청소년 전문가 1명 ○ 청소년 빅데이터 현황 파악 및 연구 세부사항 추가 확정 등을 위한 전문가 자문(2차) <ul style="list-style-type: none"> - 2020. 5. 14(목) 실시(장소: 오송역 회의실) - 자문내용: 우리나라 빅데이터 체계 구축 현황, 청소년 대상 빅데이터 온톨로지 개발 지표체계 제안 등 - 자문진: 통계청 전문가 1명, 빅데이터 관련 학계 전문가 3명 	<ul style="list-style-type: none"> ○ 빅데이터방법론 전문가 5명 ○ 청소년 핵심지표 관련 전문가 3명 ○ 청소년복지관련 전문가 3명 ○ 현장전문가 2명 등 	
5회	<ul style="list-style-type: none"> ○ 연구진행 전반에 대한 자문 <ul style="list-style-type: none"> - 빅데이터 분류를 위한 핵심지표 확정을 위한 자문(1회) - 빅데이터 온톨로지 개발을 위한 자문(4회) 		

3) 양적 연구

주요 요소	내용	비고
모집단	○ 빅데이터, 청소년, 청소년 핵심지표 관련 전문가	
표집틀	○ 학계, 현장, 공무원 30명 이상	
표본수	○ 30명 이상	
표집방법	○ 빅데이터 관련 전문가 추천을 바탕으로 눈덩이 표집	
조사시기	○ 11월~12월	

4) 질적 연구

○ 심층면접

주요 요소	내용	비고
면담참여자 섭외과정	○ 외부 공동연구원, 자문진 추천을 통해 섭외	
면담참여자 수	○ 10명 내외	
면담참여자 특성 개요	○ 학계전문가, 현장전문가, 공무원 등 각 3명 내외	
면담방식	○ 개별면접	
면담 시간 및 횟수	○ 1명당 2시간 내외	
면담자료 기록 및 녹취 여부	○ 녹취	

5) 기타 연구방법

○ 콜로키움

횟수	주요내용	참석자 범위	개최시기
1회	○ 소셜 빅데이터의 청소년 연구 활용 - 빅데이터 전문가의 방법론 등 강의 ※ 세부내용 • 소셜 빅데이터 개념과 분석 방안 • 주제 분석, 감성분석, 미래신호탐색 • 빅데이터를 활용한 머신러닝 학습 데이터 생성 • 소셜 빅데이터 연구활용 사례 등	원내 직원	2020.6.10.

○———— 제2장 이론적 배경

- 1. 빅데이터의 개요
- 2. 아동·청소년 빅데이터 관련
선행연구
- 3. 빅데이터의 활용 및 분석
- 4. 빅데이터 개인정보보호 방안

1. 빅데이터의 개요³⁾

1) 청소년 빅데이터의 개념

빅데이터(big data)는 단순 용어적으로 정의할 때 기존 데이터베이스 관리도구의 능력을 넘어서는 정형, 반정형, 비정형의 초 대량 데이터(Manyika & Chui, 2011), 또는 일상적으로 이용되는 데이터 처리 시스템, 즉 수집, 관리 및 처리 소프트웨어의 수용 한계를 넘어서는 크기의 데이터이다(Snijders, Matzat & Reips, 2012).

이런 관점에서 이미 글로벌 IT 컨설팅 기관인 Gartner(<https://www.gartner.com/>, 2020.10.1. 검색)는 빅데이터의 특징을 빠른 속도(Velocity)와 대용량(Volume) 그리고 다양성(Variety)의 3V로 규정한 바 있다. 하지만 빅데이터의 특성은 여기에 머무르지 않는다. 무엇보다 비정형 데이터의 경우 휘발성(Volatility)이 강하며 가변성(Variability)에 노출되어 있다. 정확성(Veracity)의 문제도 내재하고 있다. 따라서 빅데이터로서 활용되기 위해서는 휘발성(Volatility)이 있는 정보들 가운데 새로운 가치로 이용될 수 있는 정보가 선별되어야 하고, 정확성(Veracity)을 담보할 수 있는 가치여야 하며, 또 가변성(Variability)에 노출되어 있는 만큼

3) '제2장 이론적 배경'의 1절, 2절, 3절은 한국청소년정책연구원 성윤숙 선임연구위원이, 4절은 삼육대학교 송태민 교수가 작성하였다.

발생할 수 있는 부작용을 최소화해야 한다. 동시에 빅데이터는 그것이 활용되는 분야에 합목적적으로 상응하게 이용될 수 있어야 하고 이를 위해서는 시각화(Visualization)의 과정도 필수이다. 그러므로 실질적 차원에서 빅데이터는 빠르게 생성되는 다양한 데이터를 취합하여 양적으로만 증대된 단순거대정보가 아니라 합리적, 생산적 활용을 통해 양질의 사회적 가치로 재창출된(고수정, 2018) 거대정보를 의미한다. 여기서 합리적 생산적 활용이란 특정 사회문제와 관련된 모든 유의미한 데이터를 선별, 취합 및 생산적으로 분석하여 해당 문제의 미래지향적 해소 또는 해결을 위한 가치(Value)로서의 파생정보를 종합적으로 도출해낼 수 있는 능력 내지 기술을 뜻한다.

이상 빅데이터의 특징을 종합적으로 고려하여 ‘청소년 빅데이터’를 정의하면, 청소년 빅데이터는 ‘청소년 삶의 제반 영역에서 다양한 형태(Variety)로 빠르게(Velocity) 생성되는, 대규모(Volume) 데이터 가운데 청소년이 현실적으로 직면하는 다양한 사안의 적시성 있고 적절한 방향제시를 위해 합리적, 생산적 활용을 거쳐 유의미한 데이터로 선별, 취합, 선택적 분석 및 종합되는 모든 데이터’로 정의할 수 있다.

2) 청소년 빅데이터의 필요성

청소년 빅데이터 구축은 청소년 삶의 가치 실현이라는 영역과 삶의 질 개선이라는 영역 모두를 위해 필요하다. 먼저 긍정적인 영역에는 청소년 개인이 삶의 가치를 긍정적으로 실현하고 또 발전시킬 수 있도록 도와주기 위한 모든 작업이 포함된다. 무엇보다 청소년 개인의 측면에서 진로지도 등과 같은 생애설계나 사회활동의 기초토대가 되는 신체의 건전한 관리 등을 위해 도움을 주는 작업이 있다. 다음 부정적인 영역에는 청소년 삶의 질에 저해가 되는 제반 문제와 요소들을 해소, 해결 내지 최소화하는 작업이 포함된다.

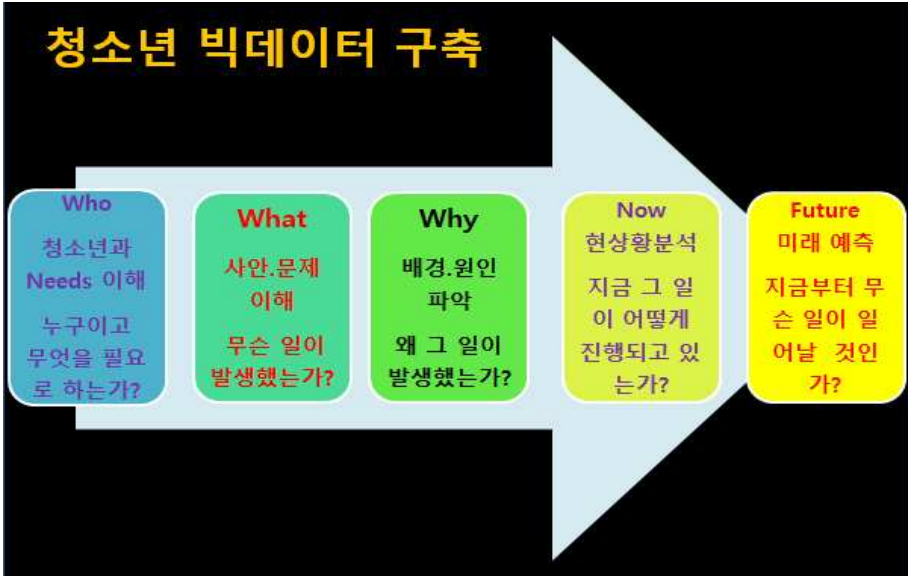


그림 II-1 청소년 빅데이터 구축의 필요성

이러한 청소년 관련 제반 영역의 효율적이고 합리적인 해소를 위해서는 기본적으로 청소년과 청소년의 니즈(needs)를 이해해야 하고 그것을 바탕으로 청소년이 당면한 제반 문제를 정확히 파악해야 한다. 청소년 빅데이터는 객관적 이해를 위한 근거제시의 도구로 기능한다. 나아가 청소년 빅데이터의 생산적 활용은 사안에 대한 원인과 현황의 정확한 분석, 종합적 진단, 진단에 기초한 증거기반 정책방향제시 및 그에 따른 미래 예측을 가능하게 해준다.

3) 국내 아동·청소년 빅데이터 구축 및 활용 현황

(1) 공공데이터 포털

공공데이터포털은 「공공데이터의 제공 및 이용 활성화에 관한 법률」 제21조(공공데이터포털의 운영)에 따라 공공기관이 생성 또는 취득하여 관리하고 있는 공공데이터를 한 곳에서 제공하는 통합 창구이다. 공공데이터활용지원센터는 공공기관의 데이터 개방과 민간의 이용 활성화를 전담 지원하는 실무 조직이며, 센터는 공공데이터 생성, 개방, 활용에 관한 전문성과 역량을 갖추고 공공과 민간의 접촉점에서 양질의 공공데이터 개방을 확대하고, 민간의 창의적 활용을 종합 지원하는 역할을 수행하고 있다(공공데이터포털, <https://www.data.go.kr/guide/guide/supportIntroduce.do>, 2020. 8. 1. 검색). 포털에서는 공공데이터를 이용할 수 있도록 파일데이터, 오픈API, 시각화 등의 방식으로 관련 정보를 제공하고 검색할 수 있다(최용환, 성운숙, 박상현, 2018 재인용).



그림 11-2 공공데이터포털 홈페이지(<https://www.data.go.kr>)에서 제공하는 청소년 통계 화면

청소년과 관련된 데이터로 파일데이터와 오픈API는 여성가족부 및 지방자치단체에서 실시하는 청소년 매체이용 및 유해환경 실태조사 등 청소년 관련 실태조사, 청소년 국제교류프로그램 정보 서비스 등 청소년 관련 프로그램 정보, 지방자치단체의 청소년활동 및 복지시설 현황, 청소년 단체 현황, 청소년 공부방 집계 현황, 청소년 공부방 현황 등의 관련 데이터를 제공한다.

(2) 정부부처

① 과학기술정보통신부

2013년 과학기술정보통신부의 전신인 미래창조과학부는 빅데이터 분석 및 활용 센터를 설립, 공공, 민간에서의 빅데이터 서비스를 구축하였으며 데이터 셋을 제공하고, 빅데이터를 활용한 청소년 복지사업 컨설팅업무를 실시한 바 있다. 그리고 현재 과학기술정보통신부와 한국정보화진흥원은 '빅데이터 기반 사회현안 해결 및 국민행복 실현'이라는 표어 아래 K-ICT 빅 데이터센터를 운영하고 있다. 이곳에서는 빅데이터와 관련한 교육 프로그램이 제공된다. 다양한 관점에서 데이터를 분석함은 물론 의미 있는 결과를 도출할 수 있는 데이터 기초프로그램부터 인공지능개발에 활용 가능한 머신러닝, 나아가 데이터 학습, 신경망 모델 이해와 성능향상 기법을 다루는 딥 러닝까지 교육받을 수 있다. 그리고 이러한 빅 데이터 사용기술은 무엇보다 청소년에 대한 긍정적 사회환경 요인을 강화하고 부정적 환경요인을 개선하거나 해소하기 위한 준거기준이 되는 지표를 제시할 수 있다.



그림 11-3 과학기술정보통신부 한국정보화진흥원 빅데이터센터(K-ICT) 홈페이지(<https://kbig.kr/portal/kbig>)

② 행정안전부

행정안전부도 '우리의 미래 빅데이터(<http://www.bigdata.go.kr>)라는 공공 분야 빅데이터 지원사업을 진행하고 있다. 목적은 다양한 데이터 융합과 분석, 그 분석을 토대로 한 주요 정책과제의 발굴 및 과학적 국가미래전략 수립이다. 세부사항으로는 먼저 빅데이터 중요성 홍보와 공공분야 빅데이터의 표준분석모델 정립 및 확산이 있다. 특히 공공 빅데이터 사업과 관련하여 2018년 청년 인턴십 커리큘럼을 마련하였고, 2019년부터 공공 빅데이터 설립을 추진하고 있다. 또 공공 빅데이터 협의회를 구성·운영하고 있으며, 2014년부터의 공공분야 빅데이터 분석과제를 사례별로 소개하고 있다.



그림 11-4 행정안전부 우리의 미래, 빅데이터 홈페이지(<http://www.bigdata.go.kr>)

그 중 특히 주목할 만한 사례는 2016년 경기도와 함께 시행한 ‘국민참여형 어린이 안전 및 교통사고 원인분석’이다. 추진배경은 우리나라 어린이 평균 교통사고 건수가 OECD 회원국들의 평균 수치보다 높다는 점(어린이 교통사고 건수가 OECD 평균 1.1명/10만명인 반면, 우리나라는 1.3명/10만명 수준)이었고, 활용 데이터는 CCTV 데이터, 교통안전시설물시스템 데이터, 그리고 119구급차량출동 시스템 데이터 등 경기도 각 기관들이 보유한 데이터들이었다.

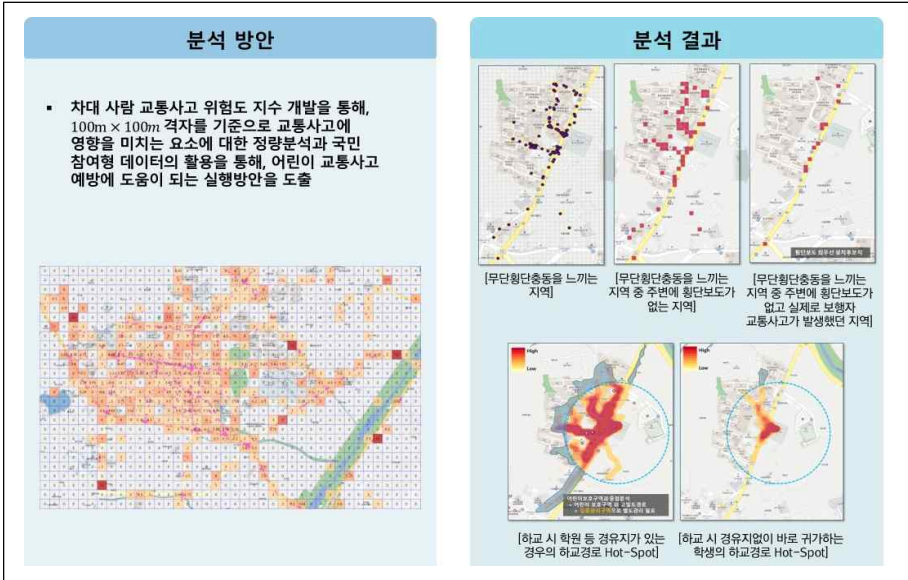


그림 11-5 행정안전부 데이터 분석 결과

이 데이터들을 분석한 결과는 교통안전시설물 입지선정을 위한 의사결정, 어린이 안전교육 자료 작성 및 안전교육, 감독 등 다양하게 활용되었으며, 학교별 맞춤형 어린이 보호구역 설정 및 어린이 통학 안전지도 콘텐츠 개발 등 다방면의 성과도 도출할 수 있었다.

③ 보건복지부

보건복지부는 건강보험데이터를 청소년 자살분석 등에 활용하고 있으며(국민건강보험공단, <https://www.nhis.or.kr/>, 2020. 8.1 검색), 사회보장정보원의 사회복지통합전산망, 국민건강보험공단 등에서 제공하는 빅데이터는 취약계층의 복지 활용에 활용되고 있다. 국민건강보험공단의 경우 2019 데이터 매직 컨퍼런스를 통해 생활안전, 건강보험 빅데이터 활용 및 건강서비스 제공을 위한 법적 타당성 검토 및 가이드라인을 제시하고 있다.

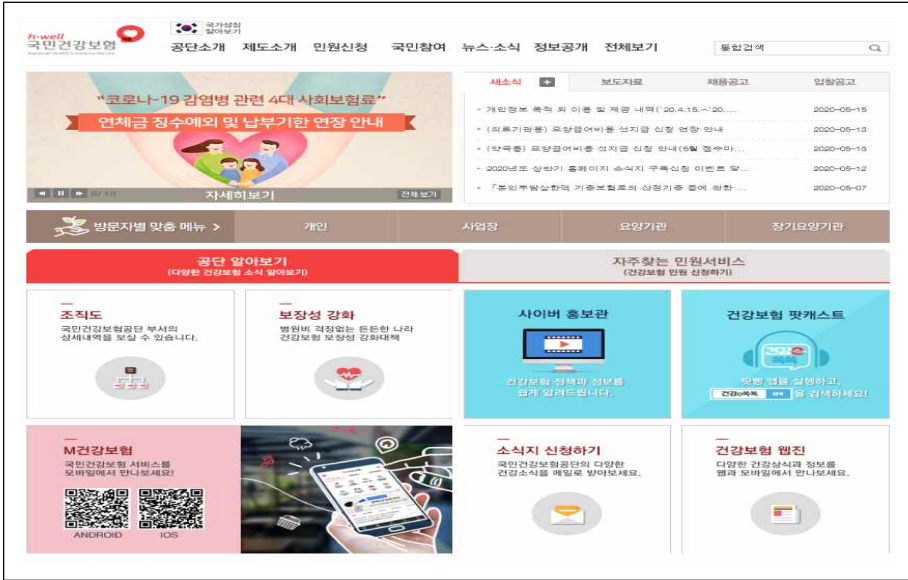


그림 II-6 보건복지부 국민건강보험공단 홈페이지(<https://www.nhis.or.kr/retrieveHomeMain.xx>)

(3) 지방자치단체

① 부산광역시 빅데이터 포털

부산광역시는 빅데이터를 지속적으로 활발하게 활용하고 있는 지방자치단체이다. 빅데이터를 활용하여 신용/부채, 서비스인구, 신용카드, 국민연금, 건강/질병/의료, 공공데이터, 보건 환경 등에 데이터시각화 활용 및 분석을 주로 사용하고 있다.



그림 11-7 부산광역시 빅데이터포털 홈페이지(<https://bigdata.busan.go.kr/index.busan>)

(4) 국책연구기관

① 한국청소년정책연구원

한국청소년정책연구원에서는 한국 아동·청소년 데이터 아카이브를 통해 다양한 패널조사와 횡단조사에 대한 데이터를 제공하고 있다. 대표적인 데이터로는 한국청소년패널조사, 한국아동·청소년패널조사, 다문화청소년패널조사, 학업중단청소년패널조사 등이 있다(<https://www.nypi.re.kr/archive>).

③ 한국행정연구원

한국행정연구원에서는 빅데이터 기반 국가전략 수립 활성화, 빅데이터 기반 갈등조기경보시스템 구축 및 운용방안 등에 대한 빅데이터 연구를 수행하고 있다.



그림 11-10 한국행정연구원 홈페이지(<https://www.kipa.re.kr/site/kipa/main.do>)

이상 공공데이터포털, 정부부처 및 지방자치단체, 국책연구기관 등에서 빅데이터 데이터셋을 제공하는 기관이나 연구를 수행하고 있는 현황을 살펴보았다. 대부분 정보통신, 기업이나 행정, 보건복지 등의 영역에서 활용되고 있으며, 아동·청소년 영역에서는 활용 부분이 적은 편이다. 청소년 빅데이터는 청소년이 건전하고 건강한 사회인으로 성장하기 위한 도구로 활용된다. 무엇보다 4차 산업혁명 시대 대한민국의 성장 동력 확보라는 측면에서도 빅데이터 관련 사업은 보다 청소년 친화적인 방향으로 진행되어야 할 것이다. 청소년 친화적인 빅데이터 연구와 사업은 체험활동, 복지, 그리고 교육 등 청소년의 긍정적 환경요인은 강화하고

미디어, 게임중독, 가정 및 학교폭력, 음주 및 흡연, 청소년 우울감 및 자살 등 부정적 환경요인은 해소 내지 감소시키는 방향으로 진행되어야 한다. 일례로 청소년의 진로선택(김지은, 백순근, 2016)이나 청소년의 생애설계에 대한 의미분석(윤은경, 박상희, 2019) 등을 할 수 있다. 이외 한국직업능력개발원 커리어 넷에서는 청소년의 진로 솔루션의 한 방향으로 빅 데이터 전문가를 소개하고 4차 산업혁명 시대 빅 데이터 분석가가 되기 위한 방안을 제시해주고 있다. 그리고 한국청소년활동진흥원(KYWA) 빅데이터센터는 청소년 체험활동 및 보호 복지 등 청소년 분야 데이터를 수집, 가공, 분석하여 다양한 분야에서 청소년활동촉진을 위한 연계방안을 모색하고 있다. 특히 청소년활동 교육온라인 프로그램이나 진로탐구 프로그램 정보 및 제공기관, 그리고 자기개발 프로그램 정보나 문화예술 프로그램 정보 및 제공기간 등 약126종에 이르는 청소년활동 참여정보 관련 빅데이터를 소장하고 있다(한국청소년활동진흥원, https://www.kywa.or.kr/management/public_10_01.jsp, 2020. 8. 1. 검색).

청소년에 대한 부정적 환경요인의 진단과 대응을 위한 청소년 빅데이터 연구와 관련해서는 한국정보화진흥원의 빅 데이터 전략연구센터가 대표적이다. 센터는 2012년 온라인 뉴스, 블로그, 카페, 트위터 및 페이스북 북에 내재된 소셜 빅 데이터를 버즈, 연관어, 영향력자 등을 중심으로 분석하여 자살 등 위기문제에 대해 청소년이 언급한 내용을 분석, 청소년 자살예방정책수립을 위한 주요정보로 활용될 수 있도록 하였다(권정은, 정지선, 2012). 특히 청소년 우울감의 경우는 학업의욕상실이나 대인관계기피 등 사회성 결핍은 물론 자살문제와도 연관되어 있는 심각한 문제이다. 이미 2008년 WHO는 21세기에 인류가 해결해야 할 핵심문제의 하나로 분류한 바 있다(WHO, https://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf, 2020. 10.1 검색). 질병관리본부 제16차(2020) 청소년건강행태조사 통계에 따르면 우울감 경험률은 ‘최근 12개월 동안 2주 내내 일상생활을 중단할 정도로 슬프거나 절망감을

느낀’ 경험에 대한 통계이다. 이와 관련하여 2008년부터 한국 청소년의 우울증상 경험률 추이를 보면 2008년(여학생 44.3%, 남학생 34.0%)부터 2015년까지(여학생 27.8%, 남학생 19.7%) 지속적인 감소추세를 보이다가 16년부터 다시 높아져 19년 현재까지(여학생 34.6%, 남학생 22.2%) 증가하고 있는 실정이다. 이러한 청소년 우울감과 같은 청소년의 부정적 환경요인의 감소 및 해소를 위해 소셜 빅데이터를 활용한 부정적 환경요인의 현황과 위험요인 분석이 가능하며(송태민, 2015) 위기관리를 위한 효율적 정보를 얻음으로, 그에 상응하는 교육 프로그램을 개발할 수 있다. 그 외 교육부와 여성가족부는 2013년 학업중단 예방 및 학교 밖 청소년 지원 방안의 틀 안에서 소셜빅데이터 분석을 통해 자살, 학업중단, 가출 등 ‘위기청소년 징후 조기경보’ 사업을 지원한 바 있다.

2. 아동·청소년 빅데이터 관련 선행연구

청소년 빅데이터 구축의 기초 작업이라 할 수 있는 청소년종합통계체계와 관련한 일련의 연구가 있었다. 먼저 김형주 외(2008)는 기존의 청소년 통계조사들과 신규통계 수요분석을 토대로 청소년통계 체제의 포괄적, 체계적 구축을 시도하였다. 또한 청소년통계체제를 효율적이고 안정적으로 취합하고 운영 및 관리하기 위하여 청소년종합통계시스템구축을 위한 정보화전략계획(ISP)의 수립을 시도하였다. 좀 더 구체적인 연구로는 김기현, 김창환(2011)이 있다. 아동·청소년 통계자료 현황 분석과 동시에 아동·청소년 핵심지표 영역을 진로, 직업정책 등 체계적으로 제시한 점이 긍정적이다.

또한 조주연(2018)은 2008년~2018년 인터넷 신문 기사를 중심으로 텍스트마이닝을 사용하여 청소년 문제의 이슈 변화추이를 분석하였다. 연구는 2008년~2018년까지 신문에 등장하는 뉴스 기사를 분석하였다. 동시에 텍스트마이닝을 활용하여 인터넷뉴스 기사의 주요이슈를 추출하고, 핵심 토픽을 도출함으로써 청소년

년 관련이슈에 내재된 패턴을 찾아보았다. 이를 통해 연구는 청소년 문제에서 두드러지는 토픽을 확인할 수 있었는데, 구체적으로 우울, 자살문제와 같은 정신 건강문제, 학교폭력, 성범죄 그리고 중독 등의 문제였다.

그 외 고려할 만한 기초자료는 아동·청소년·청년 통계관리 체계화를 위해 현황을 분석하고, 청소년정책 통계활용의 양적·질적 분석을 다룬 최용환, 성윤숙, 박상현(2018)의 연구가 있다.

다음으로 청소년 관련 개별 사안이 지니는 배경과 원인, 현재 상황, 그리고 각각의 문제를 대하는 청소년의 트렌드를 구체적으로 확인하기 위해서 빅 데이터 구축이 요구된다. 왜냐하면 그것을 기반으로 개별사안의 미래적 예측과 필요하면 조기개입이 어떻게 언제 이루어져야 할지 진단할 수 있을 것이기 때문이다. 실제 다양한 청소년 문제와 관련하여 빅 데이터를 이용한 접근이 이루어지고 있다. 먼저 청소년 삶의 질 개선의 부정적인 측면과 관련한 연구를 보면 송태민(2013)은 청소년 자살문제와 관련 비정형데이터인 소셜 빅 데이터를 분석하여 자살 검색에 관한 예측모형을 개발하고자 시도한 바 있다. 연구의 대상은 2011년 한 해 동안 수집된 소셜 빅데이터였으며, 이로부터 자살관련 토픽을 추출, 분석하여 자살 검색 예측모형을 개발하였다. 또한 노법래(2015)는 빅 데이터와 지리정보시스템(GIS)을 청소년 자살과 정신건강 서비스에 응용함으로써 정책 수요 수준과 자살률의 연계성을 확인하였다. 자살과 밀접하게 연결되어 있는 청소년 우울 문제와 관련해서는 송태민(2015)과 정혜실(2016)의 연구가 있다. 전자는 소셜 빅데이터의 데이터마이닝을 통해 청소년기의 우울 경향과 위험요소를 분석하였고, 후자는 소셜 데이터에 내재된 청소년 우울관련 정보의 추출과 분석을 위해 기능할 기본프레임으로서의 청소년 우울 온톨로지를 개발하였다. 그리고 이를 토대로 소셜 데이터 내 온톨로지 용어의 사용빈도 파악, 온톨로지의 개념과 청소년상담기록에 등장하는 개념 연계, 그리고 온톨로지 구조 및 표현력 평가를 시도하였다.

그 외 청소년 인터넷 중독문제에 대해 송태민 외(2014)의 연구가 있다. 연구는

인터넷을 통해 수집된 소셜빅데이터를 네트워크분석과 데이터마이닝의 의사결정 나무 분석기법을 활용하여 분석함으로써 청소년 인터넷 중독 위험의 예측모형을 개발하였다. 학교폭력 문제에 대해서는 한윤선 외(2019)의 학교폭력 소셜 빅데이터의 효율적 수집과 생산적 분석을 위한 온톨로지 구축 시도가 있다. 연구는 그러한 시도를 통해 학교폭력 대상자들에 대한 이해 및 학교폭력을 바라보는 사회적 관점에 대한 이해 제고를 시도하였다. 이와 관련하여 윤진수 외(2017)의 연구는 학교 폭력을 예방하기 위해 데이터 과학적 관점으로 다양한 요소의 폭행 연계 가능성을 보이고 관계되는 요소를 분석하여 폭력가해 가능성이 있는 학생을 예측하는 시도를 하였다. 나아가 사이버불링에 대한 송태민 외(2016)의 연구는 소셜 빅데이터와 구글(Google)을 기초로 한국과 미국에서 사이버불링 검색의 영향요인을 분석하였다.

마지막으로 청소년 혐오 문제를 다룬 안명숙(2018)의 연구가 있다. 연구는 소셜 데이터를 활용하여 대중의 ‘청소년 혐오’라는 신조어에 대한 담론을 분석하였다. 그리고 이를 통해 ‘청소년혐오’ 현상이 한국 사회에서 지니는 의미를 파악함으로써 청소년 정책과 청소년이론 및 상담을 위한 기초자료 제공을 시도하였다. 이상 청소년의 삶의 질에 부정적인 요소로 작동하는 다양한 문제들에 대하여 빅데이터에 기반한 분석과 이해를 위한 연구들을 살펴보았다.

다음으로 긍정적인 요소들에 대한 빅데이터 기반의 연구들을 보면 먼저 청소년 서비스 현황분석에 관한 백지연 외(2018)의 연구가 있다. 연구는 빅데이터를 활용하여 도서관에서 제공되는 다양한 청소년 대상 서비스현황의 분석, 파악 및 개선방안 제시를 시도하고 있다.

또한 생애설계문제와 관련하여 윤은경·박상희(2019)의 연구가 유의미하다. 연구는 빅데이터를 분석하여 청소년 생애 설계에 대한 의미를 도출하고, 이를 토대로 새로운 가치 도출을 마련하는데 기초자료를 제시하고자 하였다. 연구는 빅데이터의 분석을 통해 파악된 청소년의 생애설계가 ‘지원’, ‘교육’, ‘청소년’, ‘프로그램’과 유의미한 관계 하에 있음을 밝히고 있다. 구체적으로 의미연결망 분석의 중심부

에는 ‘지원’, ‘프로그램’, ‘교육’이 나타났으며, 중심성 분석에서 상호작용이 가장 높은 키워드는 ‘시작’이고, 근접중심성에서 중심성이 가장 분명하게 드러난 키워드는 ‘교육커리 관리’로 나타났다.

다음 학습 문제와 관련해서 유의미한 연구는 고수정(2018)이 있다. 연구는 빅데이터 군집 분석을 이용하여 학습 성취도를 종단적으로 예측하였는데 이를 통해 중학교의 학습 습관을 분석하면 추후 고등학교에서의 학업성취도는 물론 대학에서의 전공만족도도 예측 가능한 것으로 제시하고 있다.

마지막으로 보건복지 문제와 관련해서 송태민 외(2014)와 강희정(2015)의 연구가 있다. 송태민 외(2014)는 빅데이터의 보건복지 분야의 적용 사례를 근거로 보건복지 빅 데이터에 대한 효율적 관리 및 활용방안을 제시하고 있다. 그리고 강희정(2015)은 보건의료데이터의 정책적 활용을 위한 국가기본계획 수립을 제안하고 있다. 구체적으로 보건 의료 빅데이터의 생산적 활용을 위한 비전 수립과 보건 의료 빅데이터의 실질적 적용 가능성 파악, 그리고 활용 사업 개발과 보건 의료 빅데이터 기술적 인프라 구축전략, 나아가 보건 의료 빅데이터 거버넌스 구축 전략으로 구성되어 있다.

3. 빅데이터의 활용 및 분석

1) 비정형 빅데이터의 활용

스마트폰, 스마트TV, RFID, 센서 등의 급속한 보급과 모바일 인터넷과 소셜미디어의 확산으로 데이터량이 기하급수적으로 증가하고, 데이터의 생산·유통·소비 체계에 큰 변화가 생기면서 데이터가 경제적 자산이 될 수 있는 빅데이터 시대를 맞이하게 되었다(송태민, 2012). 세계 각국의 정부와 기업들은 빅데이터가 향후 국가와 기업의 성패를 가름할 새로운 경제적 가치의 원천이 될 것으로 기대하고 있으며 더 이코노미스트(The Economist), 가트너(Gartner), 맥킨지(McKinsey) 등은 빅데이터를 활용한 시장변동 예측과 신사업 발굴과 같은 경제적 가치창출 사례와 효과를 제시하고 있다.

특히, 빅데이터는 미래 국가 경쟁력에도 큰 영향을 미칠 것으로 기대되며 국가별로는 안전을 위협하는 글로벌 요인이나 테러, 재난재해, 질병, 위기 등에 선제적으로 대응하기 위해 우선적으로 도입하고 있다. 구글 독감예보서비스(www.google.org/flutrends)는 독감·인플루엔자 등 독감과 관련된 검색어 쿼리의 빈도를 조사하여 독감확산 조기경보체계를 제공하고 있다⁴⁾. 싱가포르는 테러 및 전염병으로 인한 불확실한 미래를 대비하기 위해 2004년부터 빅데이터를 분석·관리하는 RAHS(Risk Assessment & Horizon Scanning, 2020. 1. 30. 검색) 시스템을 구축하여 운영하고 있다. 영국은 HSC(The Foresight Horizon Scanning Centre, 2020. 1. 30. 검색)를 설립·운영하면서 비만대책 수립, 잠재적 위험관리(해안침식·기후변화), 전염병 대응 등 사회 전반의 다양한 문제에 빅데이터 기술을 활용하고 있다. EU는 대지진과 쓰나미로 인한 자연재난, 테러, 글로벌 위기, 참여와 네트워크 등 미래의 문제들에 대비하기 위한

4) 단, 구글 독감확산 조기경보체계는 2008년에 서비스를 시작하여 대략 2012년 경에 서비스를 종료하면서 빅데이터의 대표적인 실패 사례로 일컬어지는 만큼 이 부분에 대해서는 제한적으로 이해할 필요가 있으며 실패원인을 분석하여 추후 비정형 빅데이터 활용시 유의해야 할 필요가 있을 것으로 생각한다.

iKnow(Interconnect Knowledge) 프로젝트를 추진하여 세계 변화의 불확실성에 대응하고 있다. OECD는 빅데이터를 비즈니스 효율성을 제공하는 새로운 자산으로 인식하여 제15차 WPIIS 회의(OECD, 2011. 6)에서 빅데이터의 경제적 가치 측정을 의제로 채택하였다. 또한 2016년 제46회 다보스포럼에서는 '4차 산업혁명의 이해'를 핵심 주제로 선정하여 인공지능과 사물인터넷에서 생산되는 빅데이터의 '자동화와 연결성'에 기반한 분석과 활용을 강조하였다.

빅데이터는 다양성과 속도 그리고 규모의 방대함 등으로 인해 일반적인 데이터 처리 방식과는 차별화된 관리기법이 요구된다. 또한 트위터나 페이스북과 같은 소셜미디어에 남긴 정치·경제·사회·문화에 대한 메시지가 그 시대의 감성과 정서를 파악할 수 있는 원천으로 등장함에 따라 대중매체에 의해 수립된 정책의제는 이제 소셜미디어를 통해 파악할 수 있으며, 개인이 주고받은 수많은 댓글과 소셜 로그정보는 공공정책을 위한 공공재로서 진화 중에 있다(송영조, 2012). 이처럼 적지 않은 국가나 기업은 소셜 빅데이터를 분석·활용함으로써 새로운 경제적 효과를 낳고 일자리 창출을 도모하는 한편, 사회적 문제를 해결하고 미래를 예측하기 위하여 적극적으로 노력하고 있다. 기존에 실시하던 횡단적 조사나 종단적 조사 등을 대상으로 한 연구는 정해진 변인들에 대한 개인과 집단의 관계를 보는 데에는 유용하나 사이버상에서 언급된 개인별 문서(버즈: buzz)에서 논의된 관련 정보 상호 간의 연관관계를 밝히고 원인을 파악하는 데는 한계가 있다(송주영·송태민, 2014). 반면 소셜 빅데이터 분석은 복잡하게 얽혀 있는 현상 사이의 관계를 밝혀낼 수 있고, 그것을 기초로 다양한 사회적 문제에 대한 신뢰성 있는 예측을 가능하게 해준다. 단, 소셜 빅데이터 데이터의 대표성 문제, 일반화 문제는 인터넷의 광범위한 확대에도 불구하고 다소 논쟁의 여지가 있는 바 추가 논의가 필요할 것으로 보인다.

2) 비정형 빅데이터의 분석

빅데이터 분야에서 데이터 사이언티스트는 SNS 등 온라인 채널을 통해 취합된 비정형 데이터를 빠른 속도로 분석한다. 소셜미디어에서 정보를 뽑아내고 분석하는 방법은 크게 세 가지로 나눌 수 있다(송태민·송주영, 2015).

첫째, 비정형 텍스트로부터 자연어 처리 기술을 활용하여 유의미한 정보를 도출하는 텍스트마이닝(text mining)이 있다. 이는 비정형 텍스트 안에 내재된 연계성을 파악한 후 분류, 요약 등 처리하여 빅데이터에 함의된 가치 정보를 발견하는 것이다. 둘째, 소셜미디어 사용자의 의견(긍정·보통·부정 등)을 분석해내는 오피니언마이닝(opinion mining)은 주어진 텍스트 문장에 대한 자연어처리 기술과 감성분석 기술의 적용을 통해 이루어진다. 셋째, 네트워크 연결의 구조 및 강도에 대한 분석을 기초로 특정 메시지의 전달 또는 매개 경로, 나아가 영향력 및 영향범위를 추정해내는 네트워크 분석(network analysis)이 있다. 보다 구체적인 내용을 살펴보면 다음과 같다.

첫째, 해당 주제와 관련한 온라인 문서(우울)에 대해 분석 모델링을 실시하여 수집하고자 하는 대상과 범위를 설정한다. 다음, 크롤러 등 수집엔진(로봇)을 사용하여 뉴스나 블로그, 트위터 등 대상채널로부터 데이터를 취합한다. 이때 불용어(푸시킨·우울한 고백·하우리의 우울 등)를 지정하여 수집 시 오류가 발생하지 않도록 하고, 우울 관련 연관 키워드 그룹(우울·우울증)을 지정한다.

둘째, 수집한 우울 원데이터(raw data)는 연구자가 원상태로 분석하기에는 어려움이 있다. 따라서 수집한 비정형 데이터의 분류 및 정제 작업이 필요한데, 여기에 텍스트마이닝, 오피니언마이닝을 사용한다. 그리고 정제된 비정형 데이터는 버즈분석, 키워드분석 등으로 분석이 이루어진다.

셋째, 비정형 데이터를 정형 데이터로 변환한다. 우울 관련 주제 분석 사례를 살펴보면, 우울 관련 각각의 온라인 문서는 ID로 코드화해야 하고 문서 내에서

발생하는 키워드(진단·증상·예방·치료·대상 등)는 모두 빈도로 코드화해야 한다.

넷째, 사회현상과의 연관관계를 파악하려면 정형화된 데이터와 오프라인 통계 자료와의 연계가 필요하다. 이는 연계 대상 자료와 연계 가능한 식별자의 확인 후 오프라인 자료를 수집해 연계(link)할 수 있다.

다섯째, 오프라인 통계 자료와 연계된 정형화된 빅데이터를 분석할 때, 구조방정식모형은 요인 간 연관성과 시간 단위 변화의 흐름을 분석할 수 있고, 다층모형은 시간별(시간/일/월/년)·지역별 사회현상과 관계 있는 요인과의 연계성을 분석할 수 있다. 그리고 수집된 키워드를 분류하여 새로운 현상을 찾아내기 위해서는 머신러닝 분석이나 시각화를 실시할 수 있다.

4. 빅데이터 개인정보보호 방안

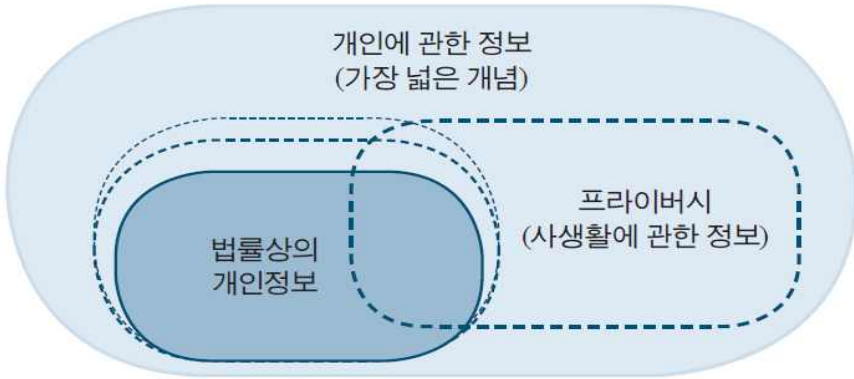
빅데이터는 크기가 방대할 뿐 아니라 처리와 분석, 그리고 활용을 모두 포함하는 개념으로 사용된다(송태민, 2012). 정부는 2011년 7월부터 행정기관과 국책연구기관이 저장하고 있는 빅데이터를 공개하였고, 기관 상호간의 공유를 통하여 국민이 활용할 수 있도록 공공데이터 공개를 추진하였다. 공공데이터는 각 기관이 자동으로 생성하고 취득하여 저장하고 있는 데이터베이스나 파일로 정부에서는 영리적 목적과 비영리적 목적에 상관없이 개발과 활용을 추진하고 있다. 특히 4차 산업혁명 시대에서는 인공지능(AI), 사물인터넷(IoT), 클라우드(Cloud), 빅데이터(Big Data) 등 지능화된 정보기술이 기존의 산업과 서비스의 활용에 융합되어 경제와 사회의 전반에 혁신적인 변화가 나타나, 과거의 산업혁명 보다 넓은 범위에서의 빠른 속도로 큰 영향을 줄 것으로 예측된다. 특히, 4차 산업혁명은 AI와 IoT 등에서 만들어지는 빅데이터의 '자동화처리와 연계화'를 기반으로 하여 처리, 분석, 활용에 중점을 두는 것으로 빅데이터 간의 공유를 통한 분석이 매우 중요하다. 어느 나라를 막론하고 빅데이터의 활용의 가장 큰 과제는 개인의 사생활

비밀 보호 및 개인정보보호이다(송태민, 2012). 이에 따라 정부는 개인정보보호에 대한 제도나 법률이 부처별로 분산되어 생기는 중복적인 규제를 제거하고 4차 산업혁명 시대에 적합한 개인과 민간 기업이 충분히 활용할 수 있는 범위를 넓히기 위해 데이터 3법을 개정하였다(2020년 1월 9일 국회 본회의 통과).

1) 개인정보, 프라이버시, 익명성의 개요

개인정보의 법률상 정의는 ‘개인정보법(제2조 제1호)’과 ‘정보통신망 이용촉진 및 정보보호 등에 관한 법률 제2조 제6호’에서 “살아 있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보(해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것을 포함한다.)를 말한다.”라고 규정하고 있다. 경제협력개발기구(Organization for Economic Cooperation and Development)는 프라이버시보호지침에서 개인적인 데이터는 알아보거나 알아볼 수 있는 개인의 모든 정보를 말한다고 정의하고 있다. 일본에서는 ‘프라이버시 보호에 관한 제도’ 상에서 말하고 있는 개인의 정보는 ‘살아있는 개인에 대한 데이터로서 해당 데이터에 포함된 이름, 태어난 년월일과 그의 개인을 판단할 수 있는 것(기타의 데이터와 쉽게 연계하여 특정의 개인을 판단할 수 있는 정보를 모두 내포한다.)’이라고 규정하고 있어 우리나라의 개인정보법에서의 표현과 유사함을 알 수 있다(송태민, 이중순, 2014).

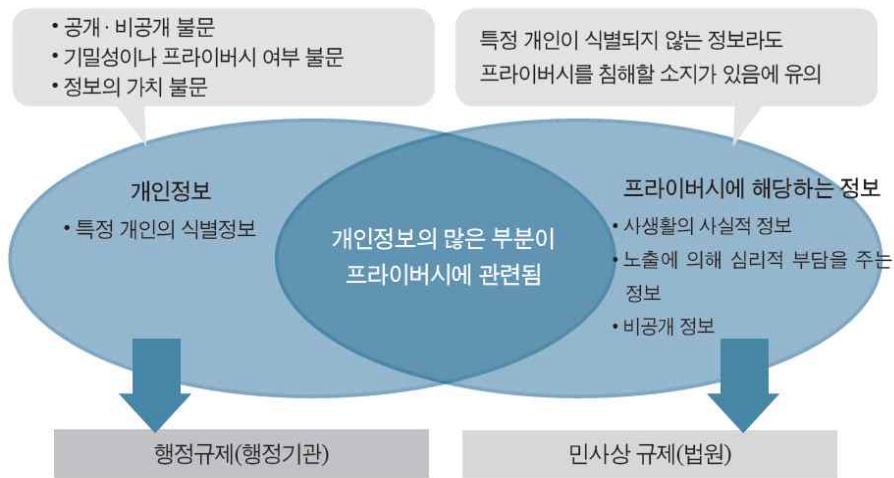
[그림 II-9]과 같이 법률과 제도에서 나타난 ‘법률상의 개인정보’와 이를 나타내는데 활용된 ‘가장 넓은 범위의 개인에 대한 정보’를 비교해 보면 개인에 관한 정보가 법률상의 개인정보를 포함하는 개념인 것을 확인 할 수 있다.



* 출처: 노무라연구소 小林慎太郎. ビッグデータ社会におけるプライバシー~「個人情報」から「プライバシー」の保護へ~第176回NRIメディアフォーラム資料. 2012. 11. 26.

그림 II-11 개인에 관한 정보와 개인정보 및 프라이버시의 관계

[그림 II-11]과 같이 개인과 관련된 정보에는 일상생활과 관련있는 프라이버시가 있다. 일반적으로 프라이버시의 의미는 개인의 사생활과 개인적인 일이나 그 정보가 다른 사람에게 표현되지 않거나, 간섭을 받지 않는 권리로 정의하고 있다. 그리고 프라이버시는 사생활(私生活)로 나타내기도 하지만 개인정보 보다 더 큰 범주의 권리를 포함한다.



* 출처: 鈴木 正朝, プライバシーの権利と個人情報保護法, 総務省マイナンバーシンポジウム資料, 2012. 11.

그림 11-12 개인정보와 프라이버시에 해당하는 정보와의 관계

익명성의 의미는 ‘개인의 원래 성명 또는 독립적 존재임을 표면적으로 나타내지 않고 숨기는 것’을 말한다. 일반적으로 익명화(anonymize)된 정보(항목)들은 개인정보보호법 상에서의 개인정보에는 포함되지 않는다는 것이다. 하지만 아직까지 익명화가 어떤 방식으로 처리/분석 되는지에 대한 정확한 설명이나 자세한 내용은 없는 현실이다(송태민, 이중순, 2014). 익명화란 구별할 수 있는 요소에 대하여 ‘1대n’의 관련성을 지키면서 ‘1:1’의 자신을 식별 할 수 있는 가능성을 제거시키는 방법이라고 판단 할 수 있다(中田響, 2007). 개인정보의 처리에 대한 한가지 방법으로 ‘가명화(pseudonymity)’가 있다.

가명화(pseudonymity)란 통상적인 구별항목을 상대적인 구별항목으로 변환하는 방법이다(송태민, 이중순, 2014). 예를 들어 <성명+구입상품내용>이 포함된 정보를 <회원의 식별자+구입상품내용>으로 교체한 것이다. 이것은 회원 식별자에 상대적으로 본인임을 알 수 있는 연결 가능성을 나타내는 익명화의 성격은 아니다.

익명화된 데이터는 구별 할 수 있는 항목들이 부족해서 다른 데이터와의 조합이 용이할 가능성이 많다고 보지 않음으로써 개인의 정보에서 빠져 나갈 수 있다. 그러나 가명화된 데이터는 특정할 수 있는 다른 사람과의 관련성이 있어서는 당연히 개인정보가 될 수 있다.

그리고 웹서비스(web service)를 사용하는 경우 개인정보를 보호하고 관리하는 입장에서 가명화를 이용하기도 한다(송태민, 이중순, 2014). 그러나 일반 사업자들은 사용자에게 지속적이고 추가적인 서비스에 대한 편리성을 위해서, 또는 사용자를 구별하고 찾아내거나 잘못된 행동을 저지할 목적으로 식별정보 등록을 강요하고 있다. 또한, 통신의 수단으로서 이메일 주소를 요구하기도 한다. 따라서 우선 식별정보를 얻게되면 비록 실명이 아닌 회원등록인 경우라도 본인을 확인할 가능성이 높은 것으로 판단하고 있다.

2) 빅데이터 활용에서의 개인정보보호방안

빅데이터의 활용에서 가장 문제가 되는 것은 개인 본인의 여러가지 저장된 정보를 수집하여 처리함으로써 발생 할 수 있는 프라이버시에 대한 침해의 문제이다. 이에 현 정부는 데이터 3법을 개정하여 규정하였다.

데이터 3법은 개인정보를 어떻게 다뤄야 하는지 규정한 3가지 법(개인정보보호법, 정보통신망법, 신용정보법)을 통칭하는 것으로 비식별 처리가 되어 있는 개인정보를 산업적 처리 목적과 연구 분석의 목적으로는 명시적으로 비동의로 활용할 수 있게 한 것이다. 따라서 데이터 3법에서는 개인에 대한 확실한 동의가 없으면 우선적으로 비식별 조치를 한 후, 연구목적으로 사용할 수 있게 하고 있다.

비식별 조치는 데이터셋에서 개개인을 알아볼 수 있는 항목에 대해 모두나 일부분을 제거하거나 교체하는 처리를 통하여 개개인을 몰라보도록 처리하는 것이다. 그리고 타 정보와 손쉽게 연계하여 본인을 알아볼 수 있는지를 체크(check)하는

‘비식별 조치 적정성 평가’를 수행한 후, 해당 데이터가 개인에 관한 정보에 속하지 않을 때만 별도의 처리없이 빅데이터 분석과 활용을 할 수 있다.

비식별 조치에 있어 가장 중요한 것은 식별자(identifiers)와 속성자(attribute values)이다(IT DAILY, 2020). 식별자는 하나 또는 연결을 통해 본인을 직접 알아볼 수 있는 항목이며, ① 고유식별정보(주민등록번호, 여권번호, 운전면허번호 등) ② 성명, ③ 상세주소, ④ 날짜 정보(생일, 기념일 등), ⑤ 전화번호, ⑥ 의료기록번호, ⑦ 계좌 및 신용카드 번호, ⑧ 자동차 및 각종 기기의 등록 번호, ⑨ 사진, ⑩ 신체 식별정보, ⑪ 이메일 및 IP, MAC 주소, ⑫ 식별코드(아이디, 직원번호, 고객번호 등) 등이 해당된다. ‘속성자’는 혼자로는 개인을 실별할 수 있는 정보는 아니지만, 타 정보와의 결합을 통하여 특정한 사람을 유추할 수 있는 데이터로, ① 신체특성, ② 개인 특성, ③ 신용 특성, ④ 가족 특성, ⑤ 전자적 특성, ⑥ 경력 특성 등이 포함된다.

현재 프라이버시 보호를 위한 구체적인 방안으로는 법과 제도에서 규정하고 있는 정보보호의 처리를 지키는 차원에서 정보를 보호하는 관리체계와 프라이버시를 보호할 수 있는 영향평가 등과 같이 시스템 상의 활용과 관리 등에 대한 여러 가지 방법이 활용되고 있다. 기술적인 면에서도 프라이버시 보호 데이터마이닝(Privacy-Preserving Data Mining, PPDM) 등 많은 방법이 개발되어 활용되고 있다(송태민, 이중순, 2014). 그러나 개인정보보호를 너무 강조하게 되면 빅데이터의 처리와 사용에 문제가 될 수 있다. 이는 서비스 공급자와 이용자가 서로 간에 여러가지 이익을 가질 수 있는 시기를 잃게 되는 것이다. 따라서 법과 제도, 그리고 기술적인 방법들을 적절하게 혼합, 활용하여 정보를 사용하는 개인이 안심하고 본인의 데이터를 제공하게 하여 빅데이터를 보다 적극적으로 사용하게 하는 일은 매우 중요하다고 할 수 있다(송태민, 이중순, 2014).

빅데이터를 적극적으로 활용하기 위해서는 다양한 매체나 기기에서 생성되는 데이터의 연계가 필요하다. 빅데이터를 연결 시키는 방법으로는 확실한 연계

(exact matching)와 통계적인 연계(statistical matching)가 있다. 확실한 연계는 고유하게 개인을 알아볼 수 있는 항목이 있을 때 사용하며, 통계적인 연계는 고유하게 개인을 알아볼 수 있는 항목이 없기 때문에 비슷한 객체를 찾아서 서로 데이터를 연계시킬 때 활용한다. 소셜 빅데이터와 정부 기관 간의 빅데이터의 결합은 시간과 관련된 항목과 지역과 관련된 항목 등을 고유한 인식정보로 하여 서로 간에 연계하는 확실한 매칭 방법을 사용할 수 있다[표 II-1].

표 II-1 빅데이터 분석기반의 위기청소년 예측 및 적시대응 기술개발 연계 사례

비정형 빅데이터	1388 정형 빅데이터	최종 연계 빅데이터
일자	일자	일자
성적	가출	성적
성	가정폭력	성
우울	학업중단	우울
경제	학교폭력	경제
질병	성	질병
외모	흡연	외모
자살위험	음주	자살위험
	자살	가출
	인터넷중독	가정폭력
	은둔형	학업중단
	비행	학교폭력
		성
		흡연
		음주
		자살
		인터넷중독
		은둔형
		비행

3) 시사점

현재 빅데이터를 활용하는 데 있어 가장 큰 과제는 개인의 프라이버시 보호 및 개인정보보호이다. 개인정보보호법 등과 관련된 제도나 법률을 자체적으로 판단하여 저장된 개인의 정보를 공공의 목적으로 사용하기 위해 다른 사람(제3자)에게 제공하는 조직은 많다. 개인정보보호를 위해 법제도적 측면에서 한국은 옵트인(Opt-in, 개인정보를 처리하기 전에 정보주체에게 먼저 동의를 받는 방식으로, 정보를 가지고 있는 개인이 동의를 하는 경우에만 개인정보를 처리할 수 있는 방식)의 개인정보보호법이 제정되어 있으며, 미국과 유럽은 옵트아웃(Opt-out, 정보를 가지고 있는 개인 동의를 얻지 않고 개인정보를 처리하고 분석하는 방법으로 정보주체가 거부 의사를 밝힌 경우에는 개인정보 처리를 즉시 중단하는 방식)의 개인정보보호법이 제정되어 있다. 따라서 이번에 마련된 데이터 3법은 비식별화 조치를 취하면 옵트아웃 방식 보다 빅데이터의 활용도는 더욱 높아질 것으로 본다. 개인정보보호법의 목적이 ‘개인정보의 수집·유출·오용·남용으로부터 사생활의 비밀 등을 보호함으로써 국민의 권리와 이익을 증진하고’로 되어 있지만, 개인에 대한 정보와 비개인에 대한 정보를 정확히 식별하기가 어렵고 상업적 목적으로 활용하는데 있어 전자적으로 연계 되는 정보가 비개인적인 데이터라도 개인의 프라이버시에 위협을 가할 수 있다. 특히 소셜미디어에 공개된 개인정보는 위조와 변조, 그리고 오용과 남용이 용이하고 비즈니스적인 목적으로 활용하기 위해서는 데이터의 수집과 저장 등에 표출될 수 있기 때문에 프라이버시와 같은 침해의 소지가 발생할 수 있다(송태민, 이중순, 2014).

개인정보보호를 강조하면 빅데이터의 사용에 문제가 될 수 있다. 빅데이터에서 개인의 무분별한 정보 유출을 막기 위해서는 특정 개인이 누구인가를 식별 하기 못하도록 처리하는 것이다. 이를 위해서는 익명화와 정보의 접근권한 및 정보의 처리방법에 대한 제한이 중요하다. 그러나 정보접근 및 정보처리에 대한 제재를 심하게 하면 정보의 사용을 확산시킬 수 없기 때문에 빅데이터의 ‘활용과 보호의 균형’에 대한 우선적인 정책이 중요할 것이다(송태민, 2013).



제3장 청소년 빅데이터 분류 체계

- 1. 청소년 빅데이터 분류 체계 기준 자료 검토
- 2. 청소년 빅데이터 분류 체계 관련 개선 의견
- 3. 통계청 아동·청소년 관련 통계 현황 조사

1. 청소년 빅데이터 분류 체계 기준 자료 검토

빅데이터 체계⁵⁾를 구축하기 위해서는 무엇보다 청소년 빅데이터의 전체 내용이 어떠한 지형을 이루고 있는지 파악할 필요가 있다. 그러나 이 작업을 하기 이전에 선행되어야 할 사항은 청소년 관련 공공영역 정형데이터, 소셜 비정형데이터를 검색하여 분류할 지표체계를 만드는 것이다. 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 있으며 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 있다. 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 있다. 국제적 추세와 한국적 상황을 모두 고려하기 위해 국제 기구의 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 부재

5) '제3장 청소년 빅데이터 분류 체계'는 한국청소년정책연구원 서정아 선임연구위원이 작성하였다.

한 영역도 파악할 필요가 있다. 본 장에서는 청소년 빅데이터 분류 체계를 도출하기 위해 UNICEF(2007)의 아동 웰빙 지표, 김기현·김창환(2011)의 아동·청소년 핵심지표, 유민상 등(2019)의 아동·청소년 삶의 질 지표를 중점적으로 살펴보도록 하겠다⁶⁾. 첫 번째로 UNICEF의 아동 웰빙 지표 체계를 살펴보면 다음과 같다.

표 III-1 UNICEF 아동 웰빙 지표 체계

범주 (demension)	구성요소 (components)	지표 (indicators)
물질적 웰빙	아동소득빈곤	• 중위소득 50% 이하 아동(0-17세) %
	박탈(deprivation)	• 11, 13, 15세 때 자신의 가족이 못 산다고 보고한 아동의 % • 6개 미만의 교육적 자산을 가지고 있다고 보고한 15세 아동의 % • 집에 책이 10권 미만이라고 응답한 15세 아동의 %
	노동	• 자녀가 있는 노동가능 연령 가구에서 부모가 고용상태가 아닌 아동의 %
건강과 안전	출생시 건강	• 1000명 중 영아 사망률 • 출생시 2.5kg 미만 영아 %
	예방접종	• 12-23개월 영아의 홍역 예방접종 %(2003년) • 12-23개월 영아의 DPT3 예방접종 %(2002년) • 12-23개월 영아의 소아마비 예방접종 %(2002년)
	아동사망률	• 19세 이하 아동 100,000명 중 사고 또는 질병 사망자수
교육적 웰빙	성취	• 15세 때 읽기능력 성취도 • 15세 때 수학능력 성취도 • 15세 때 과학능력 성취도
	참여	• 전체 15-19세 인구 중 교육기관에 풀타임 또는 파트타임으로 소속되어 있는 아동의 %
	열망	• 전체 15-19세 인구 중 교육 또는 노동에 참여하고 있지 않은 아동 % • 저숙련 노동에 종사하기를 희망하는 15세 아동의 %

6) OECD의 웰빙 지표의 경우 아동·청소년만을 대상으로 하지 않으므로 본 장에서는 제외하되 부분적으로 지표 구성에 참고하도록 하겠다.

범주 (demension)	구성요소 (components)	지표 (indicators)
또래와 가족 관계	가족구조	<ul style="list-style-type: none"> • 11, 13, 15세 아동 중 한부모 가정에서 생활하는 아동의 % • 11, 13, 15세 아동 중 계부모 가정에서 생활하는 아동의 %
	가족관계	<ul style="list-style-type: none"> • 15세 아동 중 한 주에 여러번 가족들과 함께 식사를 하는 % • 15세 아동 중 부모가 아동과 한주에 여러번 함께 시간을 보내는 %
	또래관계	<ul style="list-style-type: none"> • 11, 13, 15세 아동 중 자신의 친구가 친절하고 도움이 된다고 응답한 %
행동과 위험	위험행동	<ul style="list-style-type: none"> • 11, 13, 15세 아동 중 한주에 적어도 한번 담배를 피운 아동 % • 11, 13, 15세 아동 중 한주에 두 번 이상 술에 취한 아동 % • 15세 아동 중 지난 12개월간 대마초를 이용한 아동 % • 15-19세 여자 청소년 1000명당 청소년 임신발생 • 15세 청소년 중 성경험이 있는 청소년 % • 15세 청소년 중 가장 최근 성경험시 콘돔을 사용한 청소년 %
	폭력경험	<ul style="list-style-type: none"> • 11, 13, 15세 청소년 중 지난 12개월간 신체적인 폭력에 연루된 청소년 % • 11, 13, 15세 청소년 중 지난 2개월간 한번 이상 따돌림 당한 경험이 있는 청소년 %
	건강행동	<ul style="list-style-type: none"> • 11, 13, 15세 청소년 중 매일 과일을 먹는 청소년 % • 11, 13, 15세 청소년 중 학교가는 날 매일 아침을 먹는 청소년 • 11, 13, 15세 청소년 중 지난 한주간 1시간 이상 신체적인 운동을 한 사람의 평균 운동일수 • 13, 15세 청소년 중 BMI 지수가 과체중으로 나온 청소년 %
주관적 복지감	건강	<ul style="list-style-type: none"> • 11, 13, 15세 청소년 중 자신의 건강을 보통 또는 열악하다고 응답한 청소년 %
	개인적 웰빙	<ul style="list-style-type: none"> • 11, 13, 15세 청소년 중 생활만족도 척도가 중간 이하인 청소년 % • 15세 청소년 중 자신이 아웃사이더 또는 소외되었다고 느낀다고 응답한 청소년 % • 15세 청소년 중 자신이 어색하고 있어야 할 곳에 있지 않은 것처럼 느낀다고(out of place) 응답한 학생 % • 15세 청소년 중 외롭다고 응답한 학생 %
	학교 웰빙	<ul style="list-style-type: none"> • 11, 13, 15세 청소년 중 학교를 매우 좋다고 응답한 청소년 %

* 출처: UNICEF(2007). An overview of child well-being in rich countries. Innocenti Report Card, 7(7), 42-45.

UNICEF의 아동 웰빙 지표는 물질적 웰빙(아동소득빈곤, 박탈, 노동), 건강과 안전(출생시 건강, 예방접종, 아동사망률), 교육적 웰빙(성취, 참여, 열망), 또래와

가족 관계(가족구조, 가족관계, 또래관계), 행동과 위험(위험행동, 폭력경험, 건강 행동), 주관적 복지감(건강, 개인적 웰빙, 학교 웰빙)으로 분류되어 있음을 알 수 있다(UNICEF, 2007). 다음으로 김기현, 김창환(2011)의 아동·청소년 핵심지표 영역체계를 살펴보면 다음 표와 같다.

표 III-2 아동·청소년 핵심지표 영역체계(안)

대분류	영역	세부영역	핵심지표
Ⅰ 인구 및 가족	1-1. 인구	1-1-1 아동·청소년 인구 규모	<ul style="list-style-type: none"> • 아동·청소년 인구 수 • 아동·청소년 인구구성비(전체 대비) • 아동·청소년 연령별(5세단위) 인구수 • 아동·청소년 성별, 지역별, 인구수(구성비)
		1-1-2 아동·청소년 인구 전망	<ul style="list-style-type: none"> • 아동·청소년 인구 추계
		1-1-3 아동·청소년 인구 동태	<ul style="list-style-type: none"> • 출생아 수 • 0세의 기대여명 • 청소년 출산율 • 저체중아 출산율 • 합계출산율 • 조혼인율 • 조이혼율 • 청소년 사망률 • 청소년 사망원인
	1-2. 가족형태	1-2-1 가족형태별 아동·청소년	<ul style="list-style-type: none"> • 양부모가정 아동·청소년수 • 한부모가정 아동·청소년수 • 재혼가정 아동·청소년수 • 조손가정 아동·청소년수 • 소년소녀가정 아동·청소년 수 • 입양가정 아동·청소년수
		1-2-2 다문화 아동·청소년	<ul style="list-style-type: none"> • 다문화가정 아동·청소년 수 • 다문화가정 자녀 재학 현황 • 탈북아동·청소년 수 • 탈북아동·청소년 재학 현황
	1-3. 가족관계	1-3-1 부모자녀관계	<ul style="list-style-type: none"> • 부모자녀간 동거여부 • 부모자녀간 의사소통 • 부모자녀간 함께 하는 활동
		1-3-2 부부관계	<ul style="list-style-type: none"> • 혼인상태(초혼/재혼/이혼) • 부부간 의사소통 • 부부갈등 빈도 및 강도 • 부부간 함께 하는 활동
		1-3-3 형제자매 관계	<ul style="list-style-type: none"> • 형제자매 수 • 형제자매간 의사소통 • 형제자매간 함께 하는 활동

대분류	영역	세부영역	핵심지표
② 교육 및 역량	2-1 공교육	2-1-1 학생현황	<ul style="list-style-type: none"> • 학령인구수 • 진학률/취학률 • 학업중단률 • 대안학교 학생수 • 학교밖 청소년수
		2-1-2 학교현황	<ul style="list-style-type: none"> • 학교수/교원수 • 교원1인당 학생수 • 교육재정 규모 • 대안학교 학교수
		2-1-3 학업성취도	<ul style="list-style-type: none"> • 과목별 학업성적 • 국제학업성취도(PISA, TIMMS) • 자기주도적 학습시간
	2-2 사교육	2-2-1 사교육 규모	<ul style="list-style-type: none"> • 사교육 참여율 • 사교육비 • 학생1인당 사교육비 • 사교육시간
		2-2-2 사교육 효과	<ul style="list-style-type: none"> • 사교육효과성 • 사교육 만족도
	2-3 학교생활	2-3-1 학교생활	<ul style="list-style-type: none"> • 공부에 대한 흥미 • 학교 소속감 • 학교생활 만족도
		2-3-2 교사관계	<ul style="list-style-type: none"> • 교사-학생간 의사소통 • 교사와의 친밀도 • 교사만족도
		2-3-3 친구관계	<ul style="list-style-type: none"> • 친한 친구 수 • 친구와의 친밀도 • 친구와의 친밀도 • 동성/이성친구수
	2-4 핵심역량	2-4-1 지적 역량 (지적 도구 활용)	<ul style="list-style-type: none"> • 읽고, 쓰고 말하는 능력 • 수학적 능력 / 과학적 능력 • 정보활용능력 • 기술활용능력 • 사고력 / 창의력
		2-4-2 사회적 역량 (사회적 상호작용)	<ul style="list-style-type: none"> • 관계지향 • 협력, 협동 • 갈등관리
		2-4-3 자율적 역량 (자율적 행동)	<ul style="list-style-type: none"> • 자기관리 • 목표와 과제수립 • 자율성

대분류	영역	세부영역	핵심지표
③ 건강 및 심리 핵심지표 체계	3-1. 건강	3-1-1 신체건강	<ul style="list-style-type: none"> • 신장 및 체중 • 비만율 • 저체중율 • 체력(달리기, 멀리뛰기, 윗몸일으키기 등) • 신체건강상태(시력이상, 피부, 구강 질환 등) • 만성질환/장애여부 • 수면시간/습관 • 정기적인 운동습관 • 운동 횟수 및 시간
		3-1-2 영양섭취	<ul style="list-style-type: none"> • 영양소별 평균 섭취비율 • 식습관 • 결식율
		3-1-3 정신건강	<ul style="list-style-type: none"> • 스트레스 • 자살의도 • 우울증 • ADHD
		3-1-4 중독	<ul style="list-style-type: none"> • 인터넷 중독 • 게임 중독 • 휴대폰 중독 • 알콜 중독 • 약물 중독
	3-2. 심리	3-2-1 자기 이해	<ul style="list-style-type: none"> • 자아존중감 • 자기효능감 • 자기통제 • 자아정체감
		3-2-2 동기 및 정서	<ul style="list-style-type: none"> • 행복감 • 성취동기 • 정서장애(분노, 우울, 공격성, 불안) • 생활만족도

대분류	영역	세부영역	핵심지표
④ 보호 및 안전 핵심지표	4-1. 보호	4-1-1 유해업소, 약물 및 유해행위	<ul style="list-style-type: none"> • 유해업소 현황 • 유해업소 청소년 이용률 • 청소년 통행금지구역/제한구역 현황 • 유해업소 출입·고용 단속 현황 • 학교환경위생정화구역 현황 • 청소년 이용유해행위 단속 현황 • 유해약물 등 판매 단속 현황
		4-1-2 유해매체	<ul style="list-style-type: none"> • 영상물 심의현황 • 간행물 심의현황 • 음반 심의현황 • 청소년유해매체물 결정고시현황 • 청소년 스스로 지킴이(yr) 운영 현황 • 유해매체물 위반사범 단속현황
		4-1-3 성보호	<ul style="list-style-type: none"> • 청소년성매매 발생 및 검거 현황 • 성매매 피해자 발생 및 조치 현황 • 성매매 가해자 발생 및 조치 현황 • 신상정보 등록 및 열람대상자 현황 • 아동·청소년 대상 성범죄자 취업제한 현황
	4-2. 안전	4-2-1 학대경험	<ul style="list-style-type: none"> • 신체학대 유무 • 정서학대 유무 • 성학대 유무 • 방임 유무
		4-2-2 안전인식	<ul style="list-style-type: none"> • 안전인식 여부 • 위험물질 주의 정도 • 안전장비 착용 유무
		4-2-3 사고 발생	<ul style="list-style-type: none"> • 교통사고 발생률 • 안전사고 발생률

대분류	영역	세부영역	핵심지표
⑤ 활동 및 문화 핵심지표	5-1. 활동	5-1-1 청소년단체활동	<ul style="list-style-type: none"> • 청소년단체 회원 가입 여부 • 청소년단체활동 참여율 • 청소년단체활동 만족도
		5-1-2 동아리활동	<ul style="list-style-type: none"> • 동아리 회원 가입 여부 • 동아리활동 참여율 • 동아리활동 만족도
		5-1-3 자치활동	<ul style="list-style-type: none"> • 자치활동 임원 참여 여부 • 자치활동 참여율 • 자치활동 만족도
		5-1-4 자원봉사활동	<ul style="list-style-type: none"> • 자원봉사활동 참여율 • 자원봉사활동 시간 • 자원봉사활동 만족도
		5-1-5 청소년수련시설 활동	<ul style="list-style-type: none"> • 청소년수련시설 이용률 • 단체수련활동(캠프) 참여율 • 단체수련활동(캠프) 만족도
		5-1-6 국제교류활동	<ul style="list-style-type: none"> • 국제교류활동참여율
		5-1-7 방과후활동	<ul style="list-style-type: none"> • 방과후 나홀로 이동수 • 방과후 주로 하는 일 • 방과후 함께 있는 사람 • 방과후 학교 이용경험/만족도 • 지역아동센터 이용경험/만족도 • 방과후아카데미 이용경험/만족도
	5-2. 문화	5-2-1 견학	<ul style="list-style-type: none"> • 박물관/미술관/문화재 이용률
		5-2-2 공연/스포츠관람	<ul style="list-style-type: none"> • 영화/연극/뮤지컬/콘서트 관람률 • 스포츠경기 관람률
		5-2-3 매체	<ul style="list-style-type: none"> • TV/비디오/라디오 시청시간 • 신문/잡지 구독시간 • 인터넷/컴퓨터 이용시간 • 휴대폰/MP3 플레이어 이용시간 • 컴퓨터 게임시간
		5-2-4 독서/도서관	<ul style="list-style-type: none"> • 도서관 이용률 • 독서시간 • 독서율 • 평균독서권수
		5-2-5 여가생활	<ul style="list-style-type: none"> • 여가시간 • 여가생활만족도 • 여가활동 순위

대분류	영역	세부영역	핵심지표
⑥ 복지 및 권리 핵심지표	6-1. 복지	6-1-1 취약, 위기청소년	<ul style="list-style-type: none"> • 아동빈곤율(절대, 상대빈곤율) • 1달러 미만 생활 아동비율 • 국민기초생활수급 아동 비율 • 취약계층 아동청소년 규모 • 요보호 아동청소년 규모 • 위기 아동청소년 규모 • 가출청소년 규모 • 학업중단 청소년 규모
		6-1-2 복지시설 이용	<ul style="list-style-type: none"> • 아동·청소년복지시설 이용 경험/만족도 • 아동·청소년보호시설 이용 경험/만족도 • 청소년쉼터 이용 경험/만족도
		6-1-3 복지서비스 이용	<ul style="list-style-type: none"> • 상담서비스 이용/만족도 • 의식주 서비스 이용/만족도
		6-1-4 복지인식	<ul style="list-style-type: none"> • 복지시설에 대한 인지도 • 복지서비스에 대한 인지도
	6-2. 권리	6-2-1 권리 및 참여 의식	<ul style="list-style-type: none"> • 아동 권리에 대한 인지도 • 권리 주장 방법 • 인권에 대한 만족도
		6-2-2 학생권리	<ul style="list-style-type: none"> • 학교 규칙, 절차 참여도 • 학교의 인권보호 여부 • 학교활동 선택 정도 • 수업 선택정도
		6-2-3 인권인식	<ul style="list-style-type: none"> • 신체에 대한 인권 인식정도 • 집회/결사 인권 인식정도 • 자치활동 인권 인식정도
		6-2-4 인권교육	<ul style="list-style-type: none"> • 인권교육 경험여부 • 인권교육에 대한 만족도
		6-2-5 처벌, 징계처리	<ul style="list-style-type: none"> • 학교규칙 처벌 방법 • 처벌, 징계의 적합성 여부

대분류	영역	세부영역	핵심지표
7 비행 및 범죄 핵심지표	7-1 비행	7-1-1. 음주/흡연/약물	<ul style="list-style-type: none"> • 흡연 경험율 • 음주 경험율 • 본드 등 흡입제 사용 여부 • 유해약물 경험율
		7-1-2 학교폭력	<ul style="list-style-type: none"> • 학교폭력 피해경험 여부 • 집단따돌림 피해경험 여부 • 가해경험, 가해종류, 인식 • 가해후 처벌, 상담, 치료 여부
		7-1-3 가출	<ul style="list-style-type: none"> • 가출청소년 현황 • 가출 후 경험행동 • 가출에 대한 인식 • 가출후 연계시설, 상담 이용여부
		7-1-4. 성관련 경험	<ul style="list-style-type: none"> • 성교육 유무, 시기, 도움정도 • 성 관련 의논상대 • 성에 대한 태도 • 성관계 경험 유무 • 임신 및 낙태 경험여부
	7-2 범죄	7-2-1. 소년범죄 경험	<ul style="list-style-type: none"> • 청소년 범죄 현황 • 학생 범죄 현황 • 소년범죄/학생범죄 처리 현황 • 기소유예 처분 청소년 현황 • 청소년 보호관찰 대상 현황 • 청소년보호관찰대상자 추이 • 소년원 현황 • 소년분류심사원 현황

대분류	영역	세부영역	핵심지표
⑧ 진로 및 직업 핵심지표	8-1. 진로	8-1-1 직업 의식 및 가치관	<ul style="list-style-type: none"> • 직업의식 • 직업 가치관 • 직업에 대한 태도 • 희망직업 여부
		8-1-2 진로 준비	<ul style="list-style-type: none"> • 진로교육 활동 경험 여부/만족도 • 진로체험 활동 경험 여부/만족도 • 직업정보 취득 여부/취득 경로 • 진로/적성검사 경험 여부/만족도 • 진로 상담 경험 여부/만족도 • 진로 성숙도
	8-2. 직업	8-2-1 청소년 경제활동	<ul style="list-style-type: none"> • 청소년 경제활동인구 • 청소년 경제활동참여율 • 청소년 취업자수 • 청소년 취업자 임금 • 청소년 취업자 근로시간 • 청소년 취업률, 고용률 • 청소년 실업자수 • 청소년 실업률 • 신규졸업자 취업률 • 청소년 고용보험가입률 • 청소년 비경제활동인구 • 청소년 니트(NEET) 규모 • 청소년 고용보험 가입률 • 청소년 산재보험 가입률
		8-2-2 청소년 아르바이트	<ul style="list-style-type: none"> • 청소년 아르바이트 경험비율 • 청소년 아르바이트 현황 • 청소년 야간 및 휴일근로 인가 현황 • 연소근로자 근로조건 보호지도
		8-2-3 직업훈련 및 자격	<ul style="list-style-type: none"> • 청소년 직업훈련 현황 • 청소년 자격증 취득 현황

대분류	영역	세부영역	핵심지표	
9 정책 및 인프라 핵심지표	9-1. 정책	9-1-1 사업현황	<ul style="list-style-type: none"> • 아동급식지원현황 • 디딤돌싸앗통장 운영 현황 • 드림스타드 운영 현황 • 시설보호아동 지원 현황 • 지역아동센터 운영 현황 • 방과후 학교/방과후 교실 운영 현황 • 방과후 아카데미 운영 현황 • CYS-NET 운영 현황 • 청소년 1388 운영 현황 • 청소년 상담 실적 현황 • 수련활동인증제 실적 현황 • 청소년 문화존 운영 현황 • 청소년 동아리 지원 현황 • 청소년 국제성취포상제 현황 • 청소년 국제교류 운영 현황 	
		9-1-2 재정현황	<ul style="list-style-type: none"> • 아동청소년 예산 현황 • 청소년 육성기금 조성 및 운영 현황 	
	9-2. 인프라	9-2-1 시설현황	9-2-1 시설현황	<ul style="list-style-type: none"> • 아동·청소년복지시설 운영 현황 • 아동·청소년보호시설 운영 현황 • 청소년 쉼터 운영 현황 • 청소년 수련시설 운영 현황 • 청소년 이용시설 운영 현황
			9-2-2 단체현황	<ul style="list-style-type: none"> • 아동단체 현황 • 청소년단체 현황
		9-2-3 행정기관 현황	9-2-3 행정기관 현황	<ul style="list-style-type: none"> • 여성가족부(주무부처) 현황 • 주무부처 담당 공무원 정원 현황 • 부처별 청소년 업무 추진 현황 • 시도 및 시군구 행정조직 현황 • 정부 산하기관 현황 • 시도 청소년활동진흥센터 현황 • 시군구 청소년(상담)지원센터 현황
			9-2-4 교육 및 연구기관 현황	<ul style="list-style-type: none"> • 아동·청소년 연구기관 현황 • 아동·청소년 관련 학과 현황

* 출처: 김기현, 김장환(2011), 한국 아동·청소년 종합통계체계 구축 연구. 서울: 한국청소년정책연구원. 71-79.

김기현, 김창환(2011)의 핵심체계는 아동·청소년 관련 다양한 통계수치와 서비스, 프로그램을 총망라하고 있어 빅데이터 온톨로지 개발을 위한 청소년 지표체계를 구성하는데 참고할 바가 많을 것으로 사료된다.

세 번째로 유민상 외(2019)의 한국 아동·청소년 삶의 질 지표 체계는 사회적 배경 지표(인구영역, 사회정책 영역), 아동·청소년 삶의 질 지표(물질적 상황 및 주거환경 영역, 건강영역, 학습·발달·역량 영역, 여가·활동·참여 영역, 안전·행동 영역, 관계 영역, 주관적 웰빙 영역) 목록으로 구성되어 있으며 공공 승인 통계를 기반으로 구성된 것으로 국가차원의 공신력 있는 자료수집이 가능한 지표만으로 구성되었다는데 의의가 있다. 세부 사항은 다음과 같다. “① 사회적 배경 지표 목록: ① 인구 영역: 인구추계, 인구동태, ② 사회정책 영역: 사회정책, 인프라, ② 아동·청소년 삶의 질 지표 목록: ① 물질적 상황 및 주거환경 영역: 빈곤 및 생활수준, 주거환경, ② 건강 영역: 정신건강, 신체건강, 건강행동, ③ 학습·발달·역량 영역: 학습, 발달·역량, ④ 여가·활동·참여 영역: 여가 및 활동, 사회경제적 참여, 참여경험 및 의식, ⑤ 안전·행동 영역: 위험 및 행동, 안전, 유해환경 접촉, ⑥ 관계 영역: 가족관계, 또래관계, 지역사회, ⑦ 주관적 웰빙 영역: 삶의 만족도, 정서, 의미 있는 삶”이 그것이다.

본 고에서는 청소년관련 주요영역을 파악할 수 있는 자료로 특히 김기현, 김창환(2011)의 자료를 기준으로 청소년 전문가 조사와 통계청 공공데이터 아동청소년 관련 통계 검토 등을 거쳐 해당 지표의 세부 내용에 해당하는 공공 정형데이터와 비정형 소셜 데이터에 대한 빅데이터 분류 체계(온톨로지) 수정보완 방안을 제시하기로 한다.

7) 온톨로지는 빅데이터의 유형, 조건, 세부 개념 등을 도식화(map)하기 위한 빅데이터 분류 체계를 말하며 ‘관심 주제의 공유된 개념(shared concept)을 형식화하고(formalizing) 표현하기 위한(representing), 컴퓨터가 해석 가능한 지식 모델(computer-interpretable knowledge model)’로 정의된다(Kim, Park, Min, & Jeon, 2013).

2. 청소년 빅데이터 분류 체계 관련 개선 의견

전문가 조사 결과⁸⁾ 나타난 빅데이터 온톨로지 체계 개선을 위한 세부 방안은 다음과 같다. 우선, 빅데이터 온톨로지와 관련하여 추가 되어야 할 대분류 영역은 다음과 같이 제시될 수 있다.

진로 및 직업, 취업 및 진학, 창업, 교육·훈련 및 역량, 소비(용돈 취득 및 사용, 용돈을 주는 사람 등), 경제활동(취업, 아르바이트 등), 공동체의식(학교, 동아리, 지역사회 등), 인터넷 생활 영역(온라인콘텐츠 생산과 소비, SNS 상의 친구관계, SNS 활용 등), 자립 영역, 청소년 생활시간, 의식 및 가치(시민의식 등), 정치 참여, 사회관계망, 주거, 청소년프로그램 관련(참여도, 만족도, 지도자 만족도 등)

아동·청소년 빅데이터 온톨로지 개발과 관련하여 세부내용 개선방안 세부과제를 제시하면 다음과 같다.

8) 청소년 전문가 조사에 관한 세부 사항은 'VI장 전문가 조사' 부분을 참고하기 바란다.

① 인구 및 가족 대분류 개선 방안 세부 과제

구분	개선방안 의견		
	삭제	추가	수정보완
인구 및 가족 대분류(안) 개선방안	<ul style="list-style-type: none"> 1-3-2. 부부관계영역 조혼인율, 조이혼률 	<ul style="list-style-type: none"> 1-1-3 인구이동 1-1-3 아동청소년인구동태: 청소년의 결혼/출산인식(결혼하고 싶은지, 출산하고 싶은지) 1-2-2 다문화 아동청소년: 중도입국 청소년수* 1-2-2 다문화 아동청소년: 이주배경 아동청소년 현황 1-3 다문화 아동청소년 지표: 이주배경 유형에 따른 특성과 차이점 1-3-1 부모자녀관계: 부모와 같이 살지 않고 있는 아동청소년 1-3-1 부모의 학력 가족관계: 가족간 응집성, 부모자녀의 사소통, 조부모여부, 부모의 한부모여부, 부모양육태도, 부모자녀간 애착 지역별 아동청소년 인구동태(최소 시군구) 아동청소년 교급별 학령인구 청소년 행복 1-1-1 연령대별 인구수에서 법적 청소년 연령범위의 인구수 지표 추가(기본법상 9세~24세) 1-3-1 부모자녀관계: 부모의 자녀에 대한 폭력 여부 1-3-2 부부의 경제활동, 부부의 가사분담 1-3 가족관계: 상호작용(부모양육, 부모자녀간 갈등) 다문화청소년 부모에 대한 태도(은정적, 반항적) 진로, 부모의 진로 관여 소비수준, 소비대상 1-3-1 가족생활만족도 추가 1-3-1 부모의 사회활동(직업 유무 등) 가정경제수준 부모와 함께 하는 활동 가출유무 성과 사랑 대화시간 함께 식사하는 횟수 주거형태 	<ul style="list-style-type: none"> 0세의 기대여명을 쉬운 표현으로 수정->0세의 기대수명 20대 청소년 1인가족 전체인구를 대상으로 하는지 청소년을 대상으로 하는지 지표 명확화 필요

② 교육 및 역량 대분류 개선 방안 세부 과제

구분	개선방안 의견		
	삭제	추가	수정보완
교육 및 역량 대분류(안) 개선 방안	<ul style="list-style-type: none"> 2-2-2 사교육효과 측정의 어려움으로 삭제 필요 	<ul style="list-style-type: none"> 2-1-1 학교폭력피해물 및 가해물 2-3 학교생활: 야간자율학습 참여 여부 2-3-3 이성친구와의 친밀도 청소년 재능진단, 자기개발 지표 보완 필요 진로성숙도 대안학교 유형별 청소년수 학업성취도 정서역량 도덕적 역량(도덕적 사고, 도덕적 정서) 2-2 지역별 사교육비 2-3-1 학교폭력여부 아동청소년돌봄형태 학교유형 학습시간 2-4 진로 및 직업관련 역량 방과후활동 참여여부 2-4 미디어 및 SNS 활용관련 역량 2-3-3 친구관계를 또래관계로 치환하고 또래간 폭력여부 추가 2-3 학교생활: SNS 연결수 2-4-5 언어적 역량 교내 단체활동 참여 여부 친구관계만족도 문화예술능력, 문화감수성 사교육 분야 2-4-2 사회정서역량: 리더십 	<ul style="list-style-type: none"> 비대면수업 친구와의 친밀도 중복 OECD 교육 2030 프로젝트에서 강조한 역량체계로 전환 고려 필요 인터넷 중독은 게임중독과 유사하므로 삭제 고려

③ 건강 및 심리 대분류 개선 방안 세부 과제

구분	개선방안 의견		
	삭제	추가	수정보완
건강 및 심리 대분류(안) 개선 방안	<ul style="list-style-type: none"> 알콜중독, 약물 중독 삭제 고려 	<ul style="list-style-type: none"> 3-1-4 흡연중독 3-2-1 진로정체감 추가 경계선 장애청소년 현황 심리상담 경험 유무 도박중독 추가 자살(시도) 자주 사용하는 미디어의 종류 해당미디어를 사용하는 이유 정기적인 운동습관, 땀이 날 정도로 운동한 횟수 등 정신건강-분노(anger) 학교생활적응 3-1-5 신체적 이미지 왜곡, 성형경험, 외모관리행동(색조 화장), 다이어트 경험 등 핵심역량: 봉사활동, 인성교육, 지도력, 문화활동, 대인관계, 도덕성, 준법정신 3-1-5 보건: 병원방문(치과, 대형병원, 중소병원 등), 예방접종, 청결인식, 헌혈 등 3-2-2 동기 및 정서: 그릿(Grit) 	<ul style="list-style-type: none"> 3-1-2 결식률 구체화 필요 (예: 아침식사 결식률) 3-1-3 자살의도를 자살 및 자해 의도 3-1-4 중독-> 중독 및 과몰입(과의존)으로 변경하고 인터넷, 게임, 휴대폰에 과몰입(과의존)으로 적용 정서영역은 긍정적정서로 대체 보완 필요 3-2 심리 영역을 발달로 영역화하는 방안 고려 필요

④ 보호 및 안전 대분류 개선 방안 세부 과제

구분	개선방안 의견		
	삭제	추가	수정보완
보호 및 안전 대분류(안) 개선 방안	<ul style="list-style-type: none"> 없음 	<ul style="list-style-type: none"> 4-2 안전: 청소년시설(수련시설, 복지시설, 보호시설)에서 일어난 사고율 4-2-1 학대가해자 4-2-2 안전교육이수여부 4-2-3 사고발생->코로나 등 질병에 사고 추가 4-2 안전: 감염병 안전 항목 추가 건강: 안녕과 웰빙 미디어리터러시 아동청소년의 인터넷을 통한 유해매체노출경험과 빈도 가정에서 부모에 의한 교육 실시 여부 4-2-4 안전관련활동: 안전교육 및 조치 현황 성교육, 성문화관련단체, 현황 안전교육 자기이해: 자기조절, 충동조절 동기 및 정서: 욕구, 마음치유(힐링) 학교안전관리와 청소년활동분야 안전관리 데이터 추가 필요(시설안전 등) 학교 및 가정의 안전교육 실시 여부 	<ul style="list-style-type: none"> 4-1-1 유해업소: 유해약품 등 판매 단속 현황->유해약품 등 이용 및 판매단속 현황 4-2-1 학대경험을 유엔의 아동권리 분류 영역에 따라 폭력 및 학대 경험으로 수정보완 교통사고발생률보다는 교통사고 발생 데이터(사고일시, 사고유형, 사고위치, 사고내용 등) 4-2 인권보호에 성보호와 학대경험 포함

⑤ 활동 및 문화 대분류 개선 방안 세부 과제

구분	개선방안 의견		
	삭제	추가	수정보완
활동 및 문화 대분류(안) 개선 방안	<ul style="list-style-type: none"> 5-2 문화: MP3 플레이어 이용시간 항목 삭제 	<ul style="list-style-type: none"> • 뉴미디어, SNS를 활용한 문화 생활 • 5-1-2 동아리 종류 추가 • 5-1-7 직업체험활동경험 • 5-1 활동 SNS 및 동호회 활동 • 5-2 문화생산자로서 활동한 경험 추가 • 5-2 사회 참여 영역 추가: 정치 참여(선거), 환경운동 등 시민운동 참여, 불매운동, 시위참여 등 • 동아리활동, 봉사활동, 청소년단체활동, 수련시설 활동의 경우 참여율과 더불어 유형도 조사 • 문화활동 중 문화예술 분야에서 점차 확대되는 문화예술 교육참여 경험/참여율 또는 프로그램 현황을 추가 • 진로활동 추가 • 활동 부분에 체험 부분 추가 • 휴대폰은 단순히 이용시간만 조사할 것이 아니라 주로 어떤 용도로 활용하는지 조사 필요 • 독서의 경우 어떠한 유형의 책을 읽었는지 • 문화부분에 예술활동 추가 • 사회적 관계망을 보다 의미 있게 조사 • 청소년활동참여동기 • 청소년참여기구 현황 • 여가활동 유형 • 인증수련활동 현황, 참여현황 • 청소년활동시설 현황 • 청소년활동종사자 현황 	<ul style="list-style-type: none"> • 5-1-2 동아리활동의 경우 교내외 교외활동을 구분 • 5-1-4 자원봉사활동을 시간기부, 재능기부, 물질기부로 유형화하여 자원봉사 정신 함양 • 5-2-3 매체 상세항목과 관련하여 동영상시청시간, 커뮤니티사이트(블로그, 게임, 유튜브 사이트 등) 이용시간, SNS 이용시간, 웹툰 소설 이용시간 등 세분화 • 5-1-5 청소년 수련시설 활동에 개인 참여율 포함, 단체(개인) 수련활동(캠프) 참여율 • 방과후 주로 하는 일: 학원 및 과외보충 교육 참여 • 청소년수련시설활동은 전체적인 수정 필요: 시설유형별 통계 추가(청소년수련관, 청소년문화의 집, 공공시설/도심형 시설 현황) • 5-1-6 국제교류 활동 핵심지표를 해외자원봉사활동, 문화체험 등으로 세분화 • 5-2-4 평균독서권수->연간 평균 독서권수 • 문화활동 견학은 문화기반시설 유형(도서관, 박물관, 미술관 등)으로 유형을 구분해서 제시 • 청소년수련시설활동(5-1-5)는 현재 청소년활동진흥법 개정 이 유력시되는 만큼 수련이라는 용어가 삭제되고 새롭게 유형을 구분하고 있는 내용을 확인하여 청소년시설로 구분하여 핵심지표를 보완하고 국제교류활동(5-1-6)이란 세부영역을 포함하여 영역화 필요

⑥ 복지 및 권리 대분류 개선 방안 세부 과제

구분	개선방안 의견		
	삭제	추가	수정보완
복지 및 권리 대분류(안) 개선 방안	<ul style="list-style-type: none"> 1달러 미만 생활 아동비율 삭제 6-1-1 취약, 위기청소년 -> 아동빈곤율(절대, 상대빈곤율) 6-1-1 취약 위기청소년의 가출청소년규모는 7-1-3과 중복 	<ul style="list-style-type: none"> 6-1-1 이주배경청소년 6-1-1 핵심지표: 인터넷/게임/도박중독청소년 규모 추가 6-1-2 청소년자립지원시설 이용 추가 6-1-3 복지서비스 이용에 진로 및 직업서비스이용/만족도 추가 6-2 인권익식 중 차별의식 항목 추가 6-2 성인지 추가 권리영역에서 청소년참여활동여부(청소년운영위원회, 청소년특별회의, 참여위원회 등) 복지서비스 이용의 범위가 방대하여 위기대상별 이용 시설/서비스에 대한 세부정보 추가 필요. 그래야 복지사각지대발굴연구에 도움이 될 것으로 보임 청소년의 노동권 청소년증현황 청소년증사용현황 수영, 등산, 카누 등 청소년 액티비티 추가 6-2 양성평등인식정도 추가 권리영역에서 청소년참여활동 인지도 등 추가 문화-문화활동(연극, 영화, 오페라, 공연 등) 추가 학교밖 청소년의 권리 청소년참여에 대한 인식 가정에서의 청소년권리에 대한 보장정도 가정내 주요의사결정 참여정도 	<ul style="list-style-type: none"> 6-2 권리를 따로 대분류 영역으로 하고 NYPI의 인권실태조사 항목 참고 6-1-2 아동청소년복지시설 이용경험/만족도를 시설별로 구체화 필요 6-2-1 아동권리협약이 있으나 청소년지표는 청소년 권리로 명확하게 표현 청소년상담사에 대한 데이터 포함 여부 고려 필요 일반적인 청소년 권리 항목 강화 취약청소년에 요보호청소년 규모, 가출청소년규모, 학업중단청소년 규모 등 포함 권리의 세부영역으로 참여가 포함되어 있는데(6-2-1) 참여가 체계적인 청소년분야의 기초가 되고 있는바 이를 대분류로 구분하는 현실적 접근 필요 참여현황 영역에 가정의 의사 존중경험, 학교에서의 의사존중경험, 학교내 학생리더십 참여활동 등 다양한 경험 추가 필요

⑦ 비행 및 범죄 대분류 개선 방안 세부 과제

구분	개선방안 의견		
	삭제	추가	수정보완
비행 및 범죄 대분류(안) 개선 방안	<ul style="list-style-type: none"> • 해당사항 없음 	<ul style="list-style-type: none"> • 7-2-1 소년범죄 경험에 소년원 재입원소년범죄자수(율) 포함(재범정도 확인 목적) • 7-1-2 학교폭력이나 7-1-4 성관련 경험에 성폭력 경험 추가 • 데이트 폭력 경험 • 7-2-1 촉법소년 범죄현황 추가 필요. 최근 촉법소년 범죄에 대한 사회적 관심이 높아지고 촉법소년들이 소년법을 악용하는 사례가 빈번하게 발생하는 등 사회적 이슈로 부각되고 있음 • 7-2 범죄에 사이버범죄 가담 여부 추가 필요 • 박사방, 피싱 등과 같은 온라인상의 비행 • 범죄에 대한 청소년의 인식 정도 • 비행친구 • 기본적인 성교육 정도 • 전자담배경험률 • 7-2-1 소년범죄경험에 형사 처분대상현황 • 범죄내용 추가 필요 • 소년사법 각 단계별 처분인원 • 인권침해시 대처방법 • 피해 종류 • 피해후 감정, 상담, 치료여부 	<ul style="list-style-type: none"> • 미혼모 관련 현황은 인구 및 가족에서 다루어야 할 것 같음 • 4-2 안전한 성에 관한 사항과 7-1-4 성관련 경험 중복내용 정리 필요 • 7-1-1 음주/흡연/약물의 유해 약물 경험율은 동어 반복 • 7-1-2 학교폭력에서 집단따돌림피해경험여부는 학교폭력피해 경험여부와 중복될 수 있어 조정이 필요함 • 7-1-4 성교육 유무, 시기, 도움정도, 성관련 의논상대, 성에 대한 태도, 성관계 경험 유무는 비행 및 범죄의 항목으로 보기 어려움 • 가출청소년에 대한 지표 구성에서 가정밖 청소년을 고려한 조정 및 추가가 필요함 • 가해자 입장에서 데이터가 제공되고 있는데 피해 청소년의 입장에서 데이터가 제공될 필요가 있음 • 개인수준과 집단수준 데이터에 대한 고민 필요 • 학교 밖에서 벌어지는 일반적인 범죄피해 항목 추가 필요 • 집단따돌림 피해경험 여부는 학교폭력피해경험 여부에 포함될 수 있음 • 학생범죄현황 • 학교폭력은 외현화 문제영역으로 수정하고 그 핵심지표에는 사이버따돌림, 언어폭력 등 그 밖의 폭력과 관련한 문제를 별도의 핵심지표로 포함할 필요 있음 • 청소년의 문제와 관련한 상담경험과 상담 필요성에 대한 내용 추가 필요(상담희망분야, 상담만족도, 상담기관 접근성 등)

3. 통계청 아동·청소년 관련 통계 현황 조사

통계청 국가통계포털(<https://kosis.kr>) 주제별 승인통계에서 중앙행정기관, 지방자치단체, 공사/공단, 연구기관 승인통계 등을 중심으로 아동·청소년 관련 통계 현황을 조사한 내용 또한 온톨로지 개발의 중요한 자료로 활용되었다. 아동·청소년(9세~24세)에 관한 통계치를 포함하고 있는 기관유형별 주요 통계조사자료 중 중앙행정기관 통계 예시는 다음과 같다.

- **중앙행정기관:** 검찰청 범죄분석통계, 경찰청 경찰청 범죄통계, 고용노동부 고용형태별근로실태조사, 고용노동부 산업재해현황, 과학기술정보통신부 스마트 과의존 실태, 과학기술정보통신부 정보보호실태조사, 교육부 진로교육현황조사, 교육부 특수교육실태조사, 교육부 진로교육현황조사, 교육부 특수교육실태조사, 교육부 학생건강검사통계보고, 국토교통부 아파트주거환경통계, 국토교통부 토지소유현황, 농림축산식품부 농업법인조사, 농촌진흥청 농어업인 등에 대한 복지실태조사, 문화체육관광부 광고산업조사, 문화체육관광부 국민독서실태조사, 문화체육관광부 국민생활체육조사, 문화체육관광부 국민여가활동 조사, 문화체육관광부 국민여행조사, 문화체육관광부 국민체력실태조사, 문화체육관광부 국민문화예술활동조사, 문화체육관광부 외래관광객조사, 문화체육관광부 장애인생활체육조사, 법무부 출입국자 및 체류외국인통계, 보건복지부 국민건강영양조사, 보건복지부 국민보건의료실태통계, 보건복지부 국민생활실태조사, 보건복지부 보육실태조사, 보건복지부 보호대상아동현황보고, 보건복지부 사회복지자원봉사현황, 보건복지부 사회서비스 수요공급실태조사, 보건복지부 산후조리실태조사, 보건복지부 급성심장정지조사, 보건복지부 아동종합실태조사, 여성가족부 가족실태조사, 여성가족부 국민다문화수용성 조사, 여성가족부 성폭력안전실태조사, 여성가족부 전국 다문화가족 실태조사, 여성가족부 청소년매체이용 및 유해환경 실태조사 등



제4장 외국의 아동·청소년 빅데이터 체계 구축 사례

- 1. 미국의 아동·청소년 빅데이터 체계 구축 및 활용 사례
- 2. 영국의 아동·청소년 빅데이터 체계 구축 및 활용 사례
- 3. 호주의 아동·청소년 빅데이터 체계 구축 및 활용 사례
- 4. 외국의 아동·청소년 빅데이터 체계 구축·활용 사례 소결 및 시사점

1. 미국⁹⁾의 아동·청소년 빅데이터 체계 구축 및 활용 사례

1) 서론

최근 모바일 환경의 발전이 사람에 의한 데이터와 기계가 생산하는 데이터의 증가와 함께 정보통신분야의 대용량 데이터처리 자원의 축적을 초래함으로써 빅데이터가 전세계적 주목을 받고 있다. 이미 세계의 주요 국가들은 빅데이터의 무한한 잠재성을 놓고 치열한 기술 경쟁을 하고 있으며, 미국은 빅데이터 분석 기술 분야의 선두 국가로 다른 나라의 벤치마킹 대상이 되고 있다(배지영, 조미숙, 2016; 윤창근, 2012; 한국데이터산업진흥원, 2019).

미국은 2017년 월드와이드웹(World Wide Web, WWW)재단에서 발표한 제4차 ‘오픈데이터 글로벌 보고서(ODB Global Report Fourth Edition)’에 나타난 국가별 지표(Barometer) 순위에서, 전체 114개국 중 4위인 것으로 나타난 바 있으며, 2018년 정보화수준 IMD(Institute for Management Development) 국가경쟁력 지수에서 전년도 대비 3단계를 상승하여 1위를 기록하였으며, WEF(World Economic Forum)의 세계경쟁력 지수에서도 전년 대비 1계단 상승하여 1위를 기록하였다. 2018년도 기준으로 보면 전세계 1위의 정보화 수준을

9) ‘제4장 외국의 아동·청소년 빅데이터 체계 구축 사례’는 1절은 대구가톨릭대학교 천정웅 교수, 2절은 총신대학교 손병덕 교수, 3절은 인제대학교 류시원 교수, 4절은 한국청소년정책연구원 성윤숙 선임연구위원이 작성하였다.

갖고 있음을 알 수 있다(한국정보화진흥원, 2018). 이미 미국은 뉴욕(IBM Cloud), 캘리포니아(Google Cloud), 시애틀(Microsoft Azure) 등을 중심으로 전세계의 대표적 빅데이터 분석 플랫폼은 물론,¹⁰⁾ Amazon, Starbucks, Netflix 등 빅데이터를 사용하고 있는 수 많은 글로벌 기업들을 가지고 있다.¹¹⁾

미국 연방정부는 빅데이터 활용에 관해서 범정부적 차원의 거버넌스 체제를 구축하고 있다. 사상 최초로 정부의 데이터 정책을 총괄하는 최고 데이터과학자(Chief Data Scientist, CDS) 직위를 설치하고 백악관 과학기술정책실(Office of Science and Technology Policy: OSTP)을 중심으로 “연방 빅데이터 R&D 전략계획(The Federal BIG DATA Research and Development Strategic Plan)”을 수립하였으며, 2002년부터 연방 정부 차원의 부처와 기관들이 협력하는 IT 분야의 R&D 프로그램인 NITRO(Federal Networking and Information Technology Research and Development)를 통해 기술분야의 대규모 네트워크, 고성능 컴퓨팅 시스템, 소프트웨어, 정보관리 등의 연구개발은 물론, 신기술의 사회, 경제, 노동 등에 미치는 영향에 관한 분석 등에 관한 다양한 연구를 진행하고 있다.¹²⁾

전체적으로 다원적 정책수립과 연구개발체제 속에서 학술단체, 싱크탱크, 업계, 비영리단체, 노동조합 등의 다양한 참여자들이 커뮤니티를 형성하여 정책에 영향을 미치고 있다(황수정, 2019). 빅데이터와 관련한 최근 정책은 오바마 행정부의 오픈데이터 정책을 유지하고 있는 것으로 볼 수 있으며¹³⁾ 2019년은 인공지능과 고성능 컴퓨팅 투자를 우선순위로 제시하여 빅데이터 분야의 투자가 증가되었으며, 2020년에는 인공지능을 단독분야로 추가하여 정부 지원을 더욱 강화함으로써

10) <https://builtin.com/big-data/big-data-platform>에서 2020년 7월 10일 검색

11) <https://www.icas.com/thought-leadership/technology/10-companies-using-big-data>에서 2020년 7월 10일 검색

12) https://itlaw.wikia.org/wiki/Networking_and_Information_Technology_Research_and_Development_Program 에서 2020년 7월 10일 검색

13) <https://www.informationweek.com/big-data/ai-machine-learning/trump-white-house-launches-ai-initiative/d/d-id/1333845> 에서 2020년 7월 3일 검색

데이터 관련 산업과 업무는 더욱 확장될 것으로 전망된다(한국데이터산업진흥원, 2019; Whitehouse, 2020).

아동·청소년 관련 빅데이터 체계는 미국 빅데이터 정책의 우선순위로 나타나고 있지는 않지만, 국가전략계획과 관련하여 관심을 갖고 있음을 알 수 있는데, Data.gov 서비스를 통해 공공데이터를 공유하면서 아동·청소년관련 각종 통계와 정보자료를 가진 다양하고 방대한 데이터셋(datasets)을 제공하고 있다. 미국에서 빅데이터는 아동과 청소년분야(the child and youth field)에서 하고 있는 일에 크나큰 전망(great promise)을 제공하는 하나의 새롭고 흥미로운 프론티어(a new and exciting frontier)가 되고 있음을 알 수 있다¹⁴⁾.

2) 미국의 빅데이터 체계 현황

(1) 미국 빅데이터 추진전략

미국은 2012년 3월, 백악관의 대통령 과학기술정책국(OSTP)을 중심으로 다양한 부처가 협력하여 빅데이터 핵심 기술 개발 및 사회 각 영역에의 활용, 빅데이터 전문인력양성 등을 주요 목적으로 하는 ‘빅 데이터 R&D 이니셔티브(Big Data R&D Initiative)’를 발표하였다. 그 후 2016년에는 “연방 빅데이터 R&D 전략 계획”을 NITRO를 통해 발표하여 빅데이터 방법론, 데이터 신뢰성 확보 등을 위한 R&D 전략을 제시하였다(김수연 외, 2018). “빅 데이터 R&D 전략 계획”은 7개의 R&D 전략과 18개의 세부 과제를 담고 있는데, 미국 정부의 빅데이터 R&D 방향과 새로운 빅데이터의 기능, 이를 위해 요구되는 제반사항의 구체적 전략을 포함한다. 빅데이터 인프라 강화와 데이터 산업활성화 기반 확보, 빅데이터 인력 확충과 협력 생태계 구축 및 정부와 민간의 공동협력 강화를 강조하고 있다(김수연 외, 2018; 한국정보화진흥원, 2018).

14) <https://www.childtrends.org/blog/big-data-and-social-science-research> 에서 2020년 7월 3일 검색

아동·청소년과 관련하여 ‘빅 데이터 R&D 전략계획’에서 명시적으로 확인할 수 있는 내용은 빅데이터 분야의 지속적인 성장에 따른 교육훈련 프로그램 커리큘럼 개발과 교육자원의 배분 및 데이터를 읽고, 해석하고, 의사소통할 수 있는 능력과 새로운 지식을 만들 수 있는 능력을 갖춘 시민을 기르는 것으로 요약할 수 있다. 실제로 공익 관점을 강조한 빅데이터 커리큘럼은 여성을 포함한 소수집단의 참여를 촉진하였으며, 과학·기술·공학·수학(STEM) 분야의 생산성 증대에 기여하며,교과서, 온라인 교육과정, 실습과정, 시민과학 프로젝트 등과 같은 데이터 과학 교육 자원에 대한 접근을 제공하는 것이 필요하다고 본다(정보통신기술진흥센터, 2016).

한편, 미국은 오픈 데이터 정책의 추진과 함께 부처와 기관의 데이터를 공유·관리하는 정책을 통해 적절히 분류되고 컴퓨터로 해독할 수 있는 메타데이터(metadata) 온톨로지(ontology)를 포함하여 데이터셋(datasets)의 발전을 위해 노력하면서 응용 시스템 분야, 데이터 형식, 데이터 과학 시나리오와 관련하여 데이터들을 잘 설명하고 처리과정도 알기 쉽게 제시하는 사례 개발 연구를 강화하고 있다(정보통신기술진흥센터, 2016). 최근에는 사이버 보안 강화를 정책의 최우선 과제 중 하나로 제시하고 각 부처의 관련 예산액을 확대함으로써 데이터 관리·분석·보안 관련 시장이 함께 커질 것으로 전망하고 있다¹⁵⁾.

특히, 트럼프 행정부는 2019년 2월 12일 AI에 대한 국가 역량을 높이기 위하여 연방 정부 차원에서는 최초로 ‘미국 AI 이니셔티브(American AI Initiative)’를 추진하면서 AI 분야에서 미국의 리더십을 유지하겠다는 목적을 밝혔다.¹⁶⁾ 이 전략계획에는 연구개발투자, 인프라개방, 거버넌스 표준화, 전문인력 양성, 국제협력 등의 5대 전략을 제시하고 있는데, 인프라 개방과 관련한 배경 설명에서 데이터와 컴퓨팅 자원을 학계는 물론 산업계의 AI 연구 개발자에게 적극적으로 개방하고 지원한다는 내용을 포함하고 있다(한국정보화진흥원, 2019).

15) <https://www.whitehouse.gov/ai/> 에서 2020년 7월 10일 검색

16) <https://www.whitehouse.gov/wp-content/uploads/2018/02/ap-18-research-fy2019.pdf>.
에서 2020년 6월 13일 검색

(2) 아동·청소년 빅데이터 체계

① 공공데이터 포털의 아동·청소년정보

가. Data.gov 공공데이터 서비스 포털

미국은 오바마 행정부의 임기가 시작된 2009년부터 ‘Data.gov’ 서비스로 대표되는 공공 데이터 정책을 적극적으로 추진해 왔다. 미국 정부는 데이터 활용을 위한 소프트웨어 애플리케이션을 함께 제공하고 있는데, 이를 통해 많은 사람들이 쉽게 데이터를 활용할 수 있도록 함으로써 부가가치 창출에 기여할 수 있는 환경을 구축한 것으로 보인다(Connelly, Playford, Gayle, & Dibben, 2016; 한국데이터산업진흥원, 2019). 미국의 Data.gov는 공개데이터의 홈(the home of the U.S. Government’s open data)으로 불리고 있으며, 2020년 8월 현재 22만 4653개 이상의 데이터셋(datasets)을 확인할 수 있다.¹⁷⁾

공공데이터 data.gov의 주요 주제(topic) 분야는 농업(agriculture), 기후(climate), 생태시스템(ecosystems), 에너지(energy), 지방정부(local government), 해상(maritime), 대양(ocean), 노령자 건강(old adults health) 등으로 제시되고 있다. 데이터의 양과 주제들은 각종 데이터셋이 추가됨에 따라 주기적으로 변경되고 있으며, 각 주제별로 하위 메뉴는 각기 다르게 제시되고 있다. 예를 들면, 기후 분야는 몇 개의 하위 주제가 하위메뉴로 나타나고 있지만, 지방정부에 관한 분야는 주별, 카운티별, 도시별 데이터를 제공하는 구조의 메뉴를 갖고 있다. 해상 분야의 메뉴에는 해상운송에 관한 내용을 포함하고 있으며, 대양 분야의 하위메뉴에는 주로 미국 National Ocean Council이 제공하는 데이터셋을 중심으로 대양과 연안(coasts) 및 미국 오대호(Great Lakes)에 관한 정보를 제공하고 있다¹⁸⁾.

17) <https://www.data.gov/about> 에서 2020년 7월 3일 검색

18) <https://www.data.gov/ocean/> 에서 2020년 8월 20일 검색

표 IV-1 미국 Data.gov의 Youth Program 검색에 따른 Datasets 종류(2020년 8월)

기관	데이터 내용	종류
Department of the Interior	<ul style="list-style-type: none"> • 벤톤호수 국립야생동물피난처(Benton Lake National Wildlife Refuge) : 1996년도 연간보고서. • 나머지 83개도 위와 유사한 내용으로 다른 연도의 연차보고서 	84
City of New York	<ul style="list-style-type: none"> • 학교밖 청소년을 위한 DYCD OSY (Out-of-School Youth) — 뉴욕시의 Out-of-School Youth programs을 제공하는 기관별, 장소별 시설 현황 • 가정폭력 이니시어티브, 지표 및 요인들에 관한 보고서 — 지역의 2019년 법38호(Local Law 38 of 2019)에 의해 필수적으로 요구되는 ENDGBV프로그램의 활동과 이니시어티브들에 관한 정보 • 시설데이터베이스 — 시계획 시설의 데이터베이스(FacDB)로서 35000개 이상의 공사립 시설과 프로그램 장소에 관한 정보를 집약함 	3
State of New York	<ul style="list-style-type: none"> • 가출 및 홈리스 청소년 프로그램 (Runaway And Homeless Youth Programs) — 가출 및 홈리스 청소년을 위한 프로그램을 운영할 수 있는 승인 받은 기관에 관한 정보 제공. 	1
State of Maryland	<ul style="list-style-type: none"> • DLLR 성적향상기록 Performance Improvement Measures — GED와 NEPD이수증 수여현황 - 참가일 현재 학교에 등록된 청소년 프로그램 참가자 비율 정보. 	1
City of Bloomington	<ul style="list-style-type: none"> • 청소년프로그램 — 연가 참가자 현황 	1

* 출처: <https://catalog.data.gov/> 에서 8월 25일 검색. 발췌 번역

2020년 8월 아동 청소년과 관련한 data.gov의 정보들을 알아 보기 위해 전체 데이터 셋을 대상으로 주제어 검색을 실시하였다. 우선 ‘child’를 검색어로 했을 때는 629개가 검색되었으며, youth라는 검색어를 통해서 389개의 datasets이 검색되었다. 또한 “youth data”로는 389개의 데이터셋이 검색되었고, “youth policy”로는 21개가 제공되고 있었다. “young people”를 검색어로 하였을 때는 12개가 “youth service”로는 6개, 그리고 “youth program”으로는 90개가 검색되었다. 이는 데이터명 또는 데이터를 설명하는 주제어에 해당 검색어를 포함하고 있는 것이 모두 검색된 것으로 볼 수 있다. 그러나 각 주별, 도시별로 각기 다른

이름의 데이터셋 들이 나타나고 있으며, 연방정부 차원의 전체 내용을 포괄하는 데이터를 찾기는 쉽지 않다.¹⁹⁾

이 가운데 “청소년프로그램(youth program)”으로 검색된 데이터셋들을 보면, 조직유형(organizational types)은 연방정부 (Federal Government) (84개 데이터셋) 시정부(City Government) (4개) 주 정부는 State Government (2개)로 제시되고 있다. 기관별로는 연방정부의 내무부(Department of Interior)가 가장 많은 (84개), 뉴욕시 City of New York (3개), 뉴욕주 State of New York (1개), 메릴랜드 주, State of Maryland (1개) 및 블루밍턴시 City of Bloomington (1개) 등으로 나타나고 있다. 검색된 데이터들을 기관별로 내용을 보면, 내무부의 경우는 주로 산림감시와 관련한 연차보고서류가 많으며, 뉴욕주 등은 가출 및 홈리스 청소년프로그램 자료가 제공되고 있다. 제공된 데이터셋의 포맷 형태를 보면, PDF (84개) CSV (6개) JSON (5개) XML (3개) RDF (3개) 등으로 나타나고 있다.²⁰⁾

최근 2019년도에 Data.gov는 데이터셋의 양적 증가를 크게 보이고 있으며, 2019년 10주년을 맞이하면서 프로그램에 주요한 변화가 있었다. 이는 2019년 1월 14일에 통과된 “OPEN 정부데이터법(OPEN Government Data Act)과 관련된다²¹⁾. 이 법은 연방기관들이 반드시 관련 정보를 표준화된 기계판독 데이터 포맷(machine-readable data formats)을 사용하여 메타데이터(metadata)를 Data.gov 목록에 포함하는 것을 포함하여 온라인에 공개적으로 공표하도록 규정하고 있다. Data.gov는 새로운 법을 실행하게 됨으로써 더 많은 연방 기관들이 datasets을 포함하게 된 것이다²²⁾.

19) <https://www.data.gov/about> 에서 2020년 7월 3일 검색

20) <https://www.data.gov/about> 에서 2020년 7월 3일 검색

21) <https://www.congress.gov/bill/115th-congress/house-bill/4174/related-bills>에서 2020년 7월 20일 검색

22) <https://www.data.gov/about>에서 7월 3일 검색

나. Healthdata.gov 공공데이터 서비스 포털

미국의 공공데이터와 관련하여 보건분야의 데이터는 별도의 ‘보건관련 공공데이터 제공 사이트’(healthdata.gov)에서 서비스 하고 있다. 전체적으로 2020년 8월 현재 4531 이상의 데이터셋을 제공하고 있는데, 주제(topics) 및 데이터셋은 각기 주(states) 1847개, 건강(health) 214개, 지역사회(communitiy) 57개, 질적 수준(quality) 41개, 병원(hospital) 38개, 메디케어(medicare) 27개, 전국(national) 23개, 그리고 입원환자(inpatient) 11개 등으로 제시되어 있다. 2014년도에 실시된 선행연구에는 ‘어린이 건강(child health)’으로 27개 DB가 서비스를 제공하고 있는 것으로 제시된 바 있는데(송태민, 진달래, 박대순, 박현애, 안지영, 김정선, 2014), 2020년 8월 현재로는 각 주제별 메뉴에는 데이터셋과 데이터 제공자(developer)들에 관한 내용들이 있지만, 하위 메뉴로 아동과 청소년과 관련한 용어를 직접 사용하고 있는 것을 찾을 수 없다.

이에 따라 보다 구체적으로 아동, 청소년과 관련된 데이터셋 현황을 파악하기 위해 주제어를 활용하여 검색하였는데, 2020년 8월 기준으로, 우선, 2014년도 자료와 관련하여 child health로 검색한 결과 164개가 제시되었다. 그리고 child를 검색어로 했을 때는 172개의 Dataset이 검색되었으며, children으로는 224개가 나타났고, youth로는 94개가 결과물을 확인할 수 있었다. child와 youth를 동시에 검색했을 때는 7개가 검색되었다. 또한 child data로는 1개, youth data로는 7개, young people은 3개, youth policy로는 검색된 것이 없었으며, youth program으로는 1개가 검색되었다.²³⁾

이 중에서 94개의 데이터셋이 있는 것으로 나타난 youth로 검색한 결과를 보면, 주제는 건강(health, 17개), 주(state, 17개), 커뮤니티(communitiy, 3개), 전국(national, 2개) 등으로 나타났으며, 각 주제별 연관검색어(Tags)로 제시된 것은 모두 52개 였다. 데이터 자료의 포맷은 html(41개), csv(33개) 등 모두 15개

23) <https://healthdata.gov/> 에서 8월 3일 검색

형태로 제공되고 있었다. 자료 제공기관(publisher)들은 약물남용 및 정신건강서비스국이 53개로 가장 많았으며, 질병통제 및 예방센터가 22개로 그 다음으로 많은 것으로 나타났으며, 시카고시, 매릴랜드주, 뉴욕주 및 오클라호마주 등은 각기 1개씩 탑재되어 있었다.²⁴⁾

② 아동·청소년중심 공공데이터

가. Childstats.gov의 “미국의 아동: 전국 웰빙핵심지표” 데이터

아동과 가족 통계에 관한 연방 기관간 포럼 통계(the Federal Interagency Forum on Child and Family Statistics)라는 기구를 통해 1994년 이후 현재까지 제공되고 있는 아동에 초점을 둔 공공데이터로 “미국의 아동(America’s Children): 전국 웰빙핵심지표(Key National Indicators of Well-being)”가 있다(Federal Interagency Forum on Child and Family Statistics, 2019). 이 데이터는 연차 보고서의 형태로도 제시되지만, 검색을 통해서도 해당 내용을 확인할 수 있다. 2020년 8월 현재, 2019년도 내용을 알 수 있는데 주제영역은 크게 7가지로 가족과 사회환경(family and social environment), 경제상황(economic circumstances), 보건의료(health care), 물리적 환경과 안전(physical environment and safety), 행동문제(behavior), 교육(education), 및 건강(health) 등이며, 총 41개의 지표가 있다. 해당 지표별로 관련 정부기관에서 제공하는 가장 최근의 통계로 볼 수 있다.²⁵⁾

24) <https://healthdata.gov/> 에서 8월 3일 검색

25) https://www.childstats.gov/americaschildren/data_topics.asp에서 2020년 8월 20일 검색

나. Youth.gov의 청소년 데이터

청소년을 중심으로 하는 정부기관의 데이터로 Youth.gov가 서비스 되고 있다. youth.gov는 효과적인 청소년프로그램을 만들고, 유지하고 강화하도록 도와주는 미국 정부의 공식 홈페이지이다. 청소년통계, 재정지원 정보, 지역사회 자산을 평가하는 도구(tools)를 제공하고 지역과 연방의 자원을 찾을 수 있는 지도를 만들 수 있게 해주고(generate), 증거기반 청소년프로그램을 검색하고 최근의 청소년관련 소식(news)를 지속적으로 업데이트해준다. 이 데이터베이스에는 백악관을 비롯하여, 국무부, 보건인간서비스부, 교육부 등을 포함하여 33개 연방부서와 134개 공공기관이 참여하고 있다²⁶⁾ 이 데이터에서는 총 29개의 주제를 확인할 수 있으며, 청소년건강, 방과후 프로그램, 불링예방, 시민적 관여, 취업, 홈리스 및 가출, 소년사법, 멘토링, 긍정적 청소년개발, 자살예방, 폭력예방 등을 포함하고 있다.

다. Stopbullying.gov의 청소년 데이터

정부 차원의 공식적인 아동청소년과 관련한 대규모의 데이터들은 이밖에도 특정 주제나 분야별로 다양하게 제공되고 있다(de Angelis & Perasso, 2020). 연방정부에서는 youth.gov와 별개로 집단따돌림 이슈와 관련하여 stopbullying.gov를 운영하고 있으며, 각종 따돌림예방 연수교육, 주별 법령과 정책의 소개, 아동과 청소년 및 학부모들이 따돌림 문제에서 벗어나도록 도우는 정보 제공 등을 위해 노력하고 있다. 주로 교육부와 보건 및 인간서비스부가 공동으로 따돌림 주제에 관한 정책, 연구, 소통을 조정하는 일을 하고 있으며, 불링예방상설위원회 연방파트너(Federal Partners in Bullying Prevention Steering Committee)라는 기관간 협력체제를 통해 긴밀히 협력하고 있다. 여기에는 농업부, 국방부, 내무부, 법무부, 연방무역위원회(FTC) 및 백악관 아시아계 미국인과 태평양주민 이니셔티브도 같이 참여하고 있다²⁷⁾.

26) <https://youth.gov/federal-resources>에서 8월 20일 검색

③ 민간부문의 아동·청소년 데이터포털

가. Child Trend Data Bank

Child Trend Data Bank는 아동복지와 가족복지 및 관련 지표들을 제공하는 온라인리소스(online resource)이다. DataBank에는 거의 100개에 가까운 인구 기반통계자료(population-based measures)을 포함하고 있으며, 건강(health), 안전(safety), 아동 돌봄과 교육(child care and education) 및 긍정적 및 부정적 웰빙과 관련된 행동문제(behaviors) 등의 영역을 중심으로 관련 정보를 제공하고 있다²⁸⁾.

2020년 8월 현재 Child Trend Data Bank에서 제공하고 있는 지표의 주제들에는 아동복지(Child Welfare), 코로나 19(Covid-19), 아동(Early Childhood), ECDC, 교육(education), 가족과 양육(families and parenting), 건강(health), 히스패닉(Hispanic institute), 소년사법(Juvenile Justice), 빈곤과 불평등(Poverty and Inequality), 인종형평성(Race Equity), 사회적 정서적 발달(Social and Emotional Development), 10대 임신과 보건(Teen Pregnancy and Reproductive Health), 트라우마(Trauma), 청소년개발(Youth Development) 등 15개로 나타나고 있다.²⁹⁾

또한, 알파벳 순서로 제시되어 있는 지표의 주제들에는 집단따돌림(bullying) 문제, 청소년 투표율 등 총 66개가 제공되고 있다. 다만, 이러한 66개 지표들이 15개 주제별로 각기 다르게 구분되어 사용하고 있지는 않으며, 15개 영역별로 검색하면 관련 지표들이 중복적으로 나타나고 있다. 따라서 전체적으로 15개 주제 영역에 66개 지표체계가 활용되고 있는 것으로 볼 수 있다.³⁰⁾ Child Trends

27) <https://www.stopbullying.gov> 에서 7월 20일 검색

28) <https://www.childtrends.org/indicators?a-z>에서 2020년 7월 3일 검색

29) <https://www.childtrends.org/indicators?a-z> 에서 2020년 7월 3일 검색

30) https://www.childtrends.org/indicators?a-z_에서 2020년 7월 3일 검색

DataBank는 지속적으로 업데이트하고 있으며, 접근이 용이하고 다양한 포맷의 자료를 제공하고 있다. 복수의 관련 지표들과 관련된 대주제(broader topics)들을 중심으로 소개하는 CrossCurrents라는 데이터 브리프 시리즈(Data Brief Series)를 발간한다.

나. KIDS COUNT

KIDS COUNT 데이터는 미국의 50개 각 주별 아동의 웰빙 수준을 평가하며, 주제별, 세부지표별로 각 주별 순위를 제공하고 있는 것이 특징이다. 주로 정책가, 연구자 및 옹호자들이 보다 강화된 아동, 가족 및 지역사회의 미래를 위해 지속적으로 활용하고 있다³¹⁾. 주제는 인구구성적 특성(Demographics), 경제적 웰빙(economic well-being), 교육(education), 가족과 지역사회(family & Community), 건강(health) 및 안전과 위험행동(safety & Risky behaviors) 등 6개 영역으로 구분하고 있다.³²⁾

표 IV-2 미국 2020년 KIDS COUNT에 나타난 미국아동·청소년 추세(2020년 8월)

구분	29개 주제영역	인구	비율 (2020)	비율 (2018)	비고
경제적 웰빙	빈곤아동(Children in Poverty)	12,998,000	22%	18%	악화
	부모의 불안전 고용(Children whose parents lack secure employment)	19,579,000	33%	27%	향상
	고주거 비용 가계 생활 아동 (Children living in households with a high housing cost burden)	22,566,000	41%	31%	향상
	미재학 미고용 10대 (Teens not in school and not working)	1,186,000	9.0%	7.0%	향상

31) <https://www.aecf.org/resources/2020-kids-count-data-book/>에서 7월 25일 검색

32) <https://datacenter.kidscount.org/topics/>에서 7월 25일 검색

구분	29개 주제영역	인구	비율 (2020)	비율 (2018)	비고
교육	미재학 3-4세 아동 Young children(Ages 3 and 4) Not in School	4,215,000	52%	52%	동일
	글읽기 어려운 4학년 아동(Fourth-graders not proficient in reading)	N.A.	68%	66%	향상
	수학이 떨어지는 중2학년 (Eight-graders not proficient in Math)	N.A.	67%	67%	동일
	제때에 졸업못하는 고등학생(High school students not graduating on time)	N.A.	21%	15%	향상
건강	체중미달 영아(Low birth-weight babies)	313,752	8.1%	8.3%	악화
	건강보험 없는 아동 (children without health insurance)	4,055,000	8.0%	5.0%	향상
	10만명당 아동청소년사망율 (child and teen deaths per 100,000)	19,660	26%	25%	향상
	과체중 또는 비만 10-17세 아동청소년 (Children and teens(ages 10 to 17) who are overweight or obese)	N.A.	31%	31%	동일
가족과 지역사회	한부모 가족 아동 (Children in single-parent families)	2,398,000	34%	35%	악화
	가장이 고졸 이하인 가족 아동 (Children in families where the household head lacks a high school diploma)	9,205,000	15%	13%	향상
	최빈곤 지역 거주 아동 (Childen living in high-poverty areas)	7,717,000	13%	10%	향상
	1000명당 10대 출산율 (Teen Births per per 1,000)	179,871	34%	17%	향상

* 출처: <https://datacenter.kidscount.org>에서 2020년 8월 20일 검색

*주: N.A.는 자료 없음

이 가운데, 경제적 웰빙, 교육, 건강, 가족 및 지역사회는 각기 4개씩의 하위지표를 갖고 있으며, 전국적 지역적 추세를 알 수 있다. 인구구성의 특성으로는 인종 및 민족(race and ethnicity), 연령(age), 가족출신(family nativity)으로 구분하여 제시하고 있으며,³³⁾ 연령집단은 0 to 4, 5 to 11, 12 to 14, 15 to 17의

4개로 구분하고 전체(Total less than 18)를 제공한다³⁴⁾. <표 2>는 2020년의 KIDS COUNT에 나타난 주제별 전국적 추세를 제시한 것이다.

3) 정책 사례

(1) 개인정보 보호 등 빅데이터관련 법령·제도

미국 연방정부는 빅 데이터 활용에 관해 행정명령 형태의 가이드라인으로 범 정부적 전략을 수립하고 이를 부처별로 재설계하여 추진하는 법령체계로 운영되고 있다. 이런 의미에서 대표적 근거 제도로는 2012년에 발표된 “빅데이터 연구개발 이니셔티브(Big Data Research and Development Initiative)”와 2016년 후속 세부전략으로 제시된 빅데이터 7대 R&D 전략과 18개 세부과제를 들 수 있다(한국정보화진흥원, 2018). 또한, 정부 재정데이터의 투명성 제고를 위해 “데이터법 (Data Act of 2014: Digital Accountability and Transparency Act of 2014)”을 시행하였으며, 2018년 6월 국립과학재단에서 국가 데이터 저장 네트워크(Data Storage Network) 구축 방안과 관련 법 제정으로 데이터 개방 활성화를 위한 제도적 기술적 토대를 마련하였고, 2019년에는 “오픈 정부데이터법(OPEN Government Data Act)”을 제정하였다.³⁵⁾

그러나, 미국은 빅데이터의 기술과 활용에 따른 문제점에 대해서도 관심을 갖게 되는데, 정부가 국민의 다양한 개인 정보를 수집하고 분석하는 일에 대해 개인정보 침해와 부작용에 대한 우려의 목소리가 나오기 시작했기 때문이다(안형준, 2016). 특히, 데이터 프라이버시에 대한 보호를 요구하는 사회적 관심이 높아짐에 따라 대통령 과학기술자문위원회(PCAST)에서는 빅데이터 기술 발전에 따른 기회와 혁신을 장려하는 동시에 프라이버시 보호 방안들을 제안하였다. 그 내용은 온라인

33) <https://datacenter.kidscount.org/characteristics>에서 7월 25일 검색

34) <https://datacenter.kidscount.org/characteristics>에서 7월 25일 검색

35) <https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf>에서 2020년 8월 10일 검색

상의 자료 수집 제한권을 보장하는 ‘소비자 프라이버시 권리장정(consumer privacy bill of rights)’에 대한 개선안과 프라이버시 보호의 대상이 미국 밖의 전 세계에서든 적용되어야 한다는 선언 등을 포함하는 것이다(박준석, 2019; Whitehouse, 2016).

미국은 개별 법률에서 개인 정보의 보호와 활용을 규정하고 있는데, 전체적으로 보면, 자료 보안과 프라이버시 보호를 강화하면서 기존 행정자료의 활용이 가능한 기술 개발에 관심을 갖고 있는 것으로 볼 수 있다(한국정보화진흥원, 2018). 개인 정보의 보호와 아동·청소년과 관련한 데이터 활용에 대한 내용을 담고 있는 주요 법령들은 다음과 같다.

① OPEN 정부 데이터법(OPEN Government Data Act)

미국은 “OPEN 정부데이터법(OPEN Government Data Act)”을 통해 공공 데이터 개방을 위한 정책결정에 일반인의 참여를 보장하고 공공기관들의 개인정보 보호 계획서 제출과 지침 마련을 의무화하였으며 2019년 1월 14일 “2017년 증거기반 정책결정 구축에 관한 법률(Foundations for Evidence Based Policymaking Act of 2017)”의 한 부분으로 포함되어 입법되었다.³⁶⁾ 이 법은 데이터 개방(open)에 관한 내용을 포함하고 있으나 여기에서의 OPEN은 “개방적, 공공적, 전자적 및 필요한(Open, Public, Electronic, and Necessary)”의 의미를 가진 어휘의 머리글을 딴 것이라는 점에서 “OPEN 정부데이터법”으로 이해할 수 있을 것이다. 이 법은 개별 주나 지방자치단체의 공공데이터 개방 관련 법률, 조례, 지침, 행정명령, 정책 등에 상당한 영향을 미치고 있는데, 실제로 미국의 공공데이터인 data.gov는 이 법에 따라 하나의 정책이 아니라 입법사항이 되었다³⁷⁾. 모든 연방기관들은 data.gov 카탈로그에 포함된 메타데이터

36) <https://www.congress.gov/bill/115th-congress/house-bill/4174>에서 7월 20일 검색

37) <https://www.data.gov/meta/data-gov-at-ten-and-the-open-government-data-act/>에서 7월 20일 검색

(metadata)와 함께 해당 부서의 정보를 표준화된 “기계 판독이 가능한 데이터 포맷(standardized machine-readable data formats)”으로 반드시 온라인에 개방하도록 되었기 때문이다.

이 법의 용어정의를 보면, 메타데이터(metadata)는 “데이터에 관한 구조적 혹은 기술적 정보로 콘텐츠, 형식, 출처, 권리, 정확성, 빈도, 주기성, 입상도(granularity), 게시자 혹은 책임자, 연락처 정보, 수집방법 및 기타 기술적 내용을 포함한다”고 규정하고 있으며, “기계 판독이 가능한 데이터 포맷”은 “해당 데이터에 대해 사용되는 경우 그 본래의 의미(semantic meaning)가 손실되지 않고 인간의 개입없이 컴퓨터가 쉽게 처리할 수 있는 형태의 데이터를 의미하는 것”으로 정의하고 있다.³⁸⁾

② 공공부문의 개인정보 보호관련 입법

미국은 개인정보를 법률상 보호하고 있지만, 공공부문과 민간부문을 포괄하는 종합적인 법률은 없으며, 금융, 교육 등 각 영역별 및 주별로 개별적으로 입법하고 있다. 우선, 공공부문에서 연방정부기관이 보유하고 있는 개인정보에 관한 보호 법규들이 있다. 대표적인 것으로는 “정보공개법(Freedom of Information Act, 1966)”과 “프라이버시법(Federal Privacy Act of 1974)”을 들 수 있다.

1966년의 정보공개법은 정부 공공문서에 기록된 정보들을 공개하도록 강제하면서 프라이버시의 보호를 위한 면제조항을 포함하고 있으며, 1974년의 프라이버시법은 일반 국민이 자신들의 기록에 무슨 정보가 포함되며, 어떻게 사용되는지를 알 수 있어야 하고 일정 목적으로 수집된 정보의 다른 용도로의 사용을 금지하는 규정을 담고 있다. 특히 이 법은 전세계적으로 연방정부의 개인정보 처리 행위를 규율하는 첫 번째 국가적 입법으로 본다³⁹⁾. 공공부문의 개인정보는 또한, 그 뒤에

38) <https://www.data.gov/meta/data-gov-at-ten-and-the-open-government-data-act/>에서 7월 20일 검색

39) https://www.privacy.go.kr/pic/nation_usa.do에서 7월 20일 검색

제정된 “컴퓨터보안법(Computer Security Act, 1987)”과 “컴퓨터정보 결합 및 프라이버시보호법(Computer Matching and Privacy Protection Act, 1988)” 등에 의해서도 보호받고 있다.

③ 민간부문에 관한 개인정보 보호관련 입법

민간부문과 관련하여서는 금융, 정보통신 등 각 분야에서 필요할 때 제정된 개별법들이 있다. 대표적으로는 경제 전 분야에서 소비자 보호 및 경쟁촉진을 위한 독립적 법 집행기관으로 설치되어 있는 “연방거래위원회(FTC: Federal Trade Commision)”와 관련한 입법과 규제 사례들이 있다. 이 기관은 준 사법적 독립규제위원회로서 행정절차를 통한 시정명령을 직접 내린다⁴⁰⁾. FTC는 특정 사건으로 다수의 소비자가 피해를 입은 경우, 직접 법원에 집단소송을 제기하여 소비자 피해를 보전해 준다. 아동·청소년과 관련이 있는 FTC와 관련한 주요 법령으로는 “가족 교육법 및 프라이버시법(Family Education Rights and Privacy Act, 1974)”, “아동 온라인 프라이버시 보호법(COPPA: Child Online Privacy Protection Act, 1998)” 등이 있다.

FTC는 학생 개인 정보보호와 관련하여 미국 연방정부의 교육부와 공동으로 워크숍(Student Privacy and Ed Tech)을 개최하고, COPPA 규칙이 학교 환경에 어떻게 적용되고 “가족 교육권 및 프라이버시법(Family Education Rights and Privacy Act, FERPA)”과 어떻게 교차하는가에 대한 문제를 검토하였다(한국인터넷진흥원, 2018). COPPA 규칙에는 13세 이하 아동의 개인정보를 웹 사이트 등에서 수집할 때의 부모 동의에 대한 규정을 포함하고 있다.

40) https://www.privacy.go.kr/pic/nation_usa05.do 에서 2020년 7월 20일 검색

④ 건강보험 정보 활용 및 책임성에 관한 법률(HIPPA)

미국은 1966년 보건의료체계의 효율성을 높이고자 의료관련 행정 및 금융 자료의 전자교환을 표준화하는 법률인 “건강보험 정보 활용 및 책임성에 관한 법률(HIPAA: Health Insurance Portability and Accountability Act, 1996)”을 제정하였다. 개인이 건강관리에 관한 권리를 가지고 있음을 규정하여 개인 의료 정보를 엄격하게 보호하면서도 동시에 이름, 주소, 전화번호 등 18가지의 개인식별자가 제거된 데이터는 자유롭게 이용할 수 있도록 하는 내용을 담고 있다(한국정보화진흥원, 2018).

HIPAA는 총 5편으로 되어 있고, 제2편 F부 행정의 간소화(Title II. Substitute F - Administration Simplification)에서 개인의료정보의 보호에 관한 법률 제정의 주요 내용을 규정하고 있다⁴¹⁾. HIPAA의 개인건강정보보호와 관련한 규칙을 “프라이버시 규칙(Privacy Rule)”이라고 하고 적용대상자는 건강보험자(Health Plans), 의료공급자(의료 기관, 의료 또는 건강서비스 공급자 및 건강서비스의 공급, 금액청구, 수납 등의 업무를 하는 개인 또는 단체) 및 건강정보를 관리, 처리하는 기관인 의료정보교환소(Health Care Clearinghouse) 등으로 규정하고 있다. HIPAA의 “프라이버시 규칙”은 규제대상 또는 업무제휴자들이 보유하거나 교환하는 “개인을 식별할 수 있는 건강정보 즉, 보호건강정보(Protected Health Information:PHI)”를 보호한다. 건강정보에는 환자의 이름, 주소, 생년월일, 사회보장번호 등과 함께, 개인의 과거, 현재, 미래의 신체적, 정신적 건강 상태, 개인이 제공받은 진료기록 및 의료서비스 이용에 따라 지급한 금액 등을 포함한다⁴²⁾.

41) https://en.wikipedia.org/wiki/Health_Insurance_Portability_and_Accountability_Act에서 2020년 7월 20일 검색

42) https://en.wikipedia.org/wiki/Health_Insurance_Portability_and_Accountability_Act에서 2020년 7월 20일 검색

⑤ 가족 교육권 및 프라이버시법(FERPA)

연방법인 “가족 교육권 및 프라이버시법(Family Educational Rights and Privacy Act)”은 학생들의 개인 식별 정보 (PII)와 학생 기록부 상에 기재된 학생과 직접 관련된 모든 정보를 보호한다⁴³⁾. FERPA는 미국 교육부의 재정 지원을 받는 모든 학교에 적용되며, 아동의 교육 기록물에 대해 특정한 권리를 부여한다. 이 권리는 학생이 18세가 되거나 고등학교 이상의 학교를 다니게 되면 학생에게 이관된다. 학교, 교육구 및 주 교육부(DOE: Department of Education)는 학생 정보를 문서 형식과 전자 형식으로 보관하는데, 이러한 기록물들에는 연락처 정보, 등록 양식, 성적표, 각종 증명서, 징계 알림장 등이 포함되어 있을 수 있다. 학생의 “개인 식별 정보(PII)”는 그 학생과 관련된 모든 정보를 포함하며, 학부모, 보호자 및 자격있는 학생 (18 세 이상)은 다음과 같은 권리행사가 가능하다.⁴⁴⁾

첫째, 검사하고자 하는 기록(들)을 담당하는 학교장에게 서면 요구를 통해 교육 기록을 검사 및 검토할 수 있으며 시정할 필요가 있을 때는 기록의 변경을 요청할 수 있다. 둘째, 학교가 학생의 교육기록 정보를 공개하기 전 서면 동의서를 제공할 수 있다. 셋째, 학교의 규정 위반이 의심될 경우 주 교육부(DOE)와 미국 교육부를 고소할 수 있다. 그러나, 학교 관계자들은 합당한 교육적 이해관계를 가지거나 자신의 직업적인 책임을 수행해야 하는 경우 동의 없이도 기록에 접근할 수도 있으며 교습 및 학습 개선을 위한 데이터 분석, 학생, 교사, 학교 및 프로그램 결과의 평가, 및 해당 기록의 정확성의 확인 등과 같은 경우가 해당된다.⁴⁵⁾

43) <https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html>에서 2020년 7월 20일 검색

44) <https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html>에서 2020년 7월 20일 검색

45) <https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html>에서 2020년 7월 20일 검색

(2) 담당부처 및 민간기관

① 빅데이터 고위운영그룹과 주요 연방기관

빅데이터 관련 정책은 연방정부의 부처와 기관이 “빅데이터 고위 운영그룹(Big Data Senior Steering Group)”과 함께 프로젝트를 진행하면서 지속적으로 확대하고 있다. “빅데이터 고위 운영그룹”은 빅데이터 연구개발 조정 등과 관련한 추진기구의 성격으로 NITRD 즉, 연방 정보기술 연구개발 네트워킹(Federal Networking and IT R&D) 프로그램의 일환으로 구성된 것이다. 빅데이터 활용 효과가 높은 의료, 교육 등의 분야와 관련 있는 소관 연방부처와 기관 등이 공동으로 프로젝트에 참여하고 있으며, 국립보건원(NIH), 국립과학재단(NSF), 국방부(DoD), 국방부 산하 방위고등연구계획국(DARPA), 에너지부(DoE), 미국지질조사원(USGS), 미국항공우주국(NASA), 미국해양대기관리처(NOAA) 등을 포함한다(이상엽, 2015). 이 가운데, 국립과학재단(NSF)의 경우에는 인공지능 관련 기초 연구지원을 하고, 교통부에서 자율 및 무인시스템 추가개발을 지원하며, 국방부는 빅데이터 기반 AI 알고리즘 개발(warfighting system)을 하고 있다(김수연 외, 2018; 한국전략문제연구소, 2013).

② 지역혁신을 위한 4대 빅데이터 허브 기관

미국의 빅데이터 추진체계와 관련하여 주목할 점의 다른 하나는 데이터 과학 분야의 발전을 도모하기 위해 미국 전역에 4곳의 빅데이터 지역혁신 허브(National Network of Big Data Regional Innovation Hubs: BD Hubs)를 설립하였다는 것이다⁴⁶⁾. 미국 국가과학재단(NSF)은 2015년 11월 이후 이러한 허브구축을 통해 산업, 아카데미아 및 정부의 솔루션 관계 이해 관계자들을 연결하고 소집 및 조정하는 역할을 수행한다⁴⁷⁾. 즉, 미국 전역에 존재하는 방대한 양의

46) https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505185에서 2020년 7월 3일 검색

데이터를 효과적으로 수집, 관리, 분석하고, 관련 전문기관의 협력을 통해 다양한 국가적, 지역적 문제를 해결하고 있다.

남부(South Hub), 북동부(Northeast Hub), 중서부(Midwest Hub) 및 서부(West Hub) 등으로 구분되어 있으며, 주요 대학과 민간재단, 기업 등 250여개 이상의 기관과 단체들이 참여한다. 남부의 경우는 제조업과 헬스 케어(health care)를 중심으로 하며, 북동부는 에너지, 기후변화 및 교육관련 데이터 과학(data science) 등을 중점 분야로 하고 있다. 또한, 중서부는 농업과 식품 및 스마트 시티(smart city)와 관련한 지역 허브가 되고 있으며, 서부는 맞춤형 의료 기술과 관련한 데이터 수집의 중심 허브가 되고 있다. 이와 같이 각 허브는 담당하는 지역 뿐만 아니라 중점을 두고 있는 연구분야도 각기 상이함을 알 수 있는데, 북동부 허브의 경우 교육관련 데이터 과학을 주요 임무로 포함하고 있다는 점에서 아동·청소년 관련 역할을 갖고 있다고 볼 수 있을 것이다(이상엽, 2015).

북동부 허브(Northeast Big Data Innovation Hub)의 임무는 산업, 아카데미, 비영리기관 및 정부 간의 파트너십을 구축·강화하는 것이다⁴⁸⁾. 사회적 도전과 과학적 도전에 대해 대응하고, 경제발전을 촉진하며, 전국 빅데이터 생태계의 혁신을 가속하는 것이다(nebigdatahub.org). 특히, 청소년과 관련한 주요 활동의 하나는 북동부 학생데이터단(the Northeast Student Data Corps: NSDC)을 만든 것이다. NSDC는 북동부 허브의 공동체 개발 이니셔티브이며, 데이터 과학의 기초를 미국의 북동부 전지역 학생들에게 가르치는 역할을 한다. NSDC 커뮤니티는 교육적 접근과 자료들을 가능하도록 하는 아카데미, 산업 및 비영리 단체의 지도자들을 포함하며 지원커뮤니티를 구성한 학생들이 2년제와 4년제 대학과 소수민족 교육기관의 학생들에게 온라인으로 데이터 과학을 함께 배우면서 가르치는 일을 한다⁴⁹⁾.

47) <https://www.bigdatahubs.io/>에서 2020년 7월 3일 검색

48) <https://nebigdatahub.org/news-updates/>에서 2020년 7월 3일 검색

49) <http://nebigdatahub.org/nsdc-seedfund/>에서 2020년 7월 3일 검색

③ 연방 행정관리에산국

빅데이터와 관련한 미국 연방정부의 대표적인 부서로는 연방 행정관리에산국(OMS: Office of Management and Budget)을 들 수 있는데, 미국내 빅데이터 프로젝트를 관리하는 역할을 맡아하면서 공공데이터인 data.gov를 운영하고 있다. 미국의 대표적인 빅데이터 관련 업무를 하는 행정기관으로 UN의 빅데이터 목록(Big Data Inventory)에도 등재되어 있으며, 2020년 8월 현재 UN의 목록 상에는 약 20여개의 빅데이터 프로젝트를 관리하고 있는 것으로 나타나고 있다. 빅데이터 소스들 뿐만아니라 빅데이터 기술과 관련된 것이나, 공식 통계 또는 지속가능개발 지표(SDG)와 관련된 것을 중심으로 제시되어 있으나, 아동·청소년에 관한 명시적 내용을 포함하고 있는 프로젝트는 나타나고 있지 않다⁵⁰⁾.

미국 연방행정관리에산국(OMS)에서 관리하는 공공데이터 플랫폼인 data.gov는 2019년 5월에 10주년을 맞이하고 2020년 5월 이후 11년째 운영하고 있는데, 2009년 5월에 US General Service Administration(GSA)가 47개의 데이터셋으로 출발하였으며, 2020년 8월 현재는 연방 기관, 주, 카운티 및 시를 포함하는 23만 이상의 데이터셋으로 성장하였다. Data.gov는 전세계의 다른 국가와 도시 및 여타의 개방 정부 데이터 카탈로그들이 2009년 이후 자체적인 개방 정부 데이터 사이트를 운영하게 하는 대표적인 사례가 되었다⁵¹⁾. Data.gov는 데이터셋의 접근성을 높이기 위해 노력하고 있으며, 주요 사용자들은 일반 대중, 학생, 연구자, 언론인 및 기업인들이며, 해를 거듭할수록 그 활용도가 증대하고 있다. 연간 약 2천만 페이지 뷰(page view)에 이르는 것으로 알려진다⁵²⁾.

50) <https://unstats.un.org/bigdata/inventory/?selectTitle=&selectCountry=United+States&selectOrganization=U.S.+Office+of+Management+and+Budget&selectSource=&selectstatArea=&selectsdgGoal=>에서 2020년 8월 10일 검색

51) <https://www.data.gov/meta/data-gov-at-ten-and-the-open-government-data-act/>에서 7월 3일 검색

52) <https://www.data.gov/meta/data-gov-at-ten-and-the-open-government-data-act/>에서 7월 3일 검색

④ 아동과 가족 통계에 관한 연방 기관간 포럼

아동과 가족 통계에 관한 연방 기관간 포럼(the Federal Interagency Forum on Child and Family Statistics(the Forum)은 행정명령(Executive Order) No. 13045에 의해 1997년 4월에 아동과 가족에 관한 연방 데이터를 수집하고 보고하는데 조정과 공동협력을 강화하기 위하여 설립되었다. 2019년 현재 소비자 생산안전위원회(Consumer Product Safety Commission), 농업부(Department of Agriculture), 경제연구서비스부(Economic Research Service), 상무부(Department of Commerce), 미국센서스국(U.S. Census Bureau), 국방부(Department of Defense), 교육부 등 21개 부처가 참여하고 있다⁵³⁾.

이 포럼에 참여한 연방기관의 실무연구진에서 미국의 아동과 가족 웰빙과 관련된 상황(conditions)과 경향(trends)에 관한 데이터를 수집하고 분석, 제공하는 연방정부의 노력을 조정, 공동협력 및 통합하는 것을 지원하고 있으며 데이터 플랫폼으로 Childstats.gov를 운영하고 있다.⁵⁴⁾ 또한, 이 포럼에서는 미국의 아동(America's Children): 전국 웰빙 핵심지표(Key National Indicators of Well-being)를 운영하고 있다. 이는 1994년 처음 시작되었으며, 2019년 기준으로 7개 영역 총 41개 지표를 업데이트하였다. 포럼에 참여하고 있는 기관들이 제공하는 데이터들은 미국 지역사회 서베이(American Community Survey), 시민권 데이터 통계(Civil Rights Data Collection), 인구서베이(Current Population Survey) 등의 전국적 서베이결과와 “공기의 질(Air Quality System)”, “안전식수정보시스템(Safe Drinking Water Information System)”, “청소년 위험행동감시시스템(Youth Risk Behavior Surveillance System)” 등과 같은 시스템을 포함하여 20개 데이터들을 활용하여 제공하고 있다.⁵⁵⁾

53) <https://www.childstats.gov/americaschildren/members.asp>에서 2020년 8월 20일 검색

54) <https://www.childstats.gov/>에서 2020년 8월 20일 검색

55) <https://www.childstats.gov/pubs/>에서 2020년 8월 20일 검색

⑤ 민간기관 KIDS COUNT Data Center

KIDS COUNT data center는 애니 케시 재단(Annie E. Casey Foundation)에서 운영하는 하나의 프로젝트로 시작된 것이며, 현재는 이를 통해 KIDS COUNT라는 아동과 가족에 관한 가장 중요한(a premier) 데이터 소스(source of data)를 제공하고 있다⁵⁶⁾. 재단의 기능 중 하나는 매년 KIDS COUNT Data Book이라는 종합 리포트를 발간하여 미국 아동의 웰빙을 평가하는 것이다. Data book에 제시된 지표들은 Data Center에서 개발하였다. 주목할 것은 데이터들을 미국의 각 주별, 주제별, 특징별로 검색할 수 있다는 점이다. 1993년부터 매년 데이터를 생산하여 왔으며, 아동의 웰빙수준에 관한 최근 데이터를 제공하며 각 주별로 순위를 매기고 2010년부터의 다년간 추세를 비교하는 통계데이터를 찾을 수 있도록 운영하고 있다. 주로 정책가, 연구자 및 옹호자들이 보다 강화된 아동, 가족 및 지역사회의 미래를 위해 지속적으로 활용하고 있다⁵⁷⁾.

(3) 아동·청소년 빅데이터의 활용

빅 데이터 활용 영역은 무궁무진하다. 특히, 소셜 빅데이터와 같은 비정형데이터는 정부의 정책과 새로운 기술에 대한 미래신호를 사전에 예측하여 대비할 수 있으며, 조사를 통한 기존 정보수집체계의 한계를 보완할 수 있는 새로운 자료 수집 방법으로도 활용할 수 있다(송태민, 이기호, 진달래, 천미경, 서동철, 박현애, 2016). 미국의 경우 각 연방정부기관들의 빅데이터 활용 프로젝트와 협업적 프로그램에 대한 지속적 지원이 이루어지고 있으며, 공공과 민간부문의 정형·비정형 빅데이터를 활용하는 다양한 사례를 찾아볼 수 있다. 특히, 교육빅데이터 활용과 빅데이터 연구를 통한 학습참여 증진과 취약청소년 지원과 관련한 사례 및 PSYCKES를 통한 청소년자살예방 대응체제 구축과 관련한 사례는 주목할 필요가 있다.

56) <https://datacenter.kidscount.org/>에서 8월 20일 검색

57) <https://www.aecf.org/resources/2020-kids-count-data-book/>에서 7월 25일 검색

① 청소년학업과 관련한 교육빅데이터 활용

아동·청소년과 관련한 미국의 빅데이터 정책사례의 가장 대표적인 것은 교육빅데이터(educational big data)에서 찾을 수 있다. 교육빅데이터는 교육과 관련한 대규모 자료를 의미하며 활용가치는 주로 ① 교육정책 및 교육과정에 대한 의사결정의 기준(국가와 지역), ② 학교 교사 수준에서는 학력 미달 학생 지도 및 교수 내용과 방법 개선의 준거(학교와 교사), ③ 개별화 학습 및 자기 진단의 준거(학생과 학부모) 등으로 활용될 수 있다고 본다(이은경, 박도영, 최인봉, 2014).

실제로 학습을 원활하게 지원하고, 교수자를 지원하는 기반기술로서 빅데이터의 활용 노력들이 학습지원이라는 큰 틀에서 지속되고 있다(권영옥, 2014). 미국에서는 2012년 2월 이후 미국 전역에 모든 교과서를 각종 PC 등의 디지털 기기에서 이용할 수 있는 디지털 교과서로 교체한 바 있으며, 2017년 9월 트럼프 행정부에서도 대통령지침을 통해 교육부에 초·중·고 학습의 컴퓨터교육을 강화하도록 지시하여, 디지털 교과서, 무료 전자 교과서, 무료 온라인 강좌 등을 확산하고 멀티미디어 자료를 활용한 교육 콘텐츠 이용을 강화하고 있다(한국정보화진흥원, 2018).

교육빅데이터는 특히 대학교육을 통해 그 활용이 증대하고 있는데, 학업성적관리, 소셜네트워크 데이터의 교육적 활용 및 모바일 또는 정보통신 기술을 활용한 새로운 형태의 수업방법 등과 관련하여 빅데이터가 활용되고 있는 사례를 다양하게 찾아볼 수 있다. 데이터 마이닝 기법을 이용한 맞춤형 교육서비스의 제공 사례는 물론, 소셜 네트워크 서비스의 데이터를 활용하여, 개인 성향에 맞는 적절한 과외활동(extracurricular activities)과 커뮤니티를 추천하는 등으로 학업 이외의 학교생활에 활용하고 있는 것이다(이은경 외, 2014; Harari, Muller, Aung, & Rentfrow, 2017).

② 빅데이터 활용 연구를 통한 학습참여 증진과 취약청소년 지원

빅데이터는 학생들의 학습참여에 영향을 주는 중요한 역할을 하고 있는 사례를 제시하는 연구들이 있다(Watson & Christensen, 2017). 빅데이터를 통해 교사 중심의 일방적인 수업이 아닌 학생들의 능동적인 행동적(behavioral), 정서적(emotional), 인지적(cognitive) 진정한 참여를 촉진할 수 있다고 본다(Harris, 2008; 임희준, 2020 재인용). 빅데이터는 또한 학교내 취약 청소년(vulnerable youth)과 학교밖 청소년을 위해서도 활용되고 있다. 학교행정가들에게 가용한 정보는 전통적으로 행동문제 기록이나 학업성취와 같은 변수에 한정되어 왔지만, 빅데이터의 정보들은 취약 청소년들이 직면하는 따돌림(bullying) 문제와 학교안전 문제에 관한 것으로 확대하고 있다(Giusine, Fraboni, de Angelis, & Pietrantonio, 2019). 빅데이터의 수집과 분석에 관한 기회가 발전함에 따라 이전에는 활용할 수 없었거나 접근이 어려웠던 정보를 분석함으로써 새로운 통찰력을 가질 기회를 갖는 것이다.

또한, 인구조사 결과 분석을 활용하여 Child Trends Data Bank 등이 제시한 데이터를 통해 관련 정책을 수립하거나, 학교밖 청소년에 대한 보다 체계적이고 실증적인 분석을 하는 데에도 빅데이터는 의미를 갖는다고 본다(백혜정, 송미경, 신정민, 2015). 빅데이터 연구로 청소년 우울증과 관련한 임상실무지침을 개발하고, Medicine, Cochrane, PsychoInfo 등과 같은 데이터 베이스를 통해 청소년 우울증에 대한 총체적인 임상지식을 집약하여 기분장애(mood disorder) 온톨로지를 개발한 사례가 있다(정혜실, 2016). 빅데이터를 활용한 연구 정보들은 내면적 문제와 안전문제에 취약한 학생들을 포함하여 보다 안전한 환경을 만들려고 노력하는 일에 기여할 수 있는 다양한 이점들을 제공함으로써 아동·청소년의 성공, 건강, 행복을 위해 긍정적으로 활용되고 있다(Guntuku, Yaden, Kern, Ungar, & Eichstaedt, 2017).

③ 청소년위험과 범죄예측을 통한 대응체제 구축 사례

청소년 건강정보와 소셜 데이터 등을 활용한 빅데이터 분석을 통해 청소년들의 위험징후 패턴을 도출하고 실시간 대응하는 것이 가능하며 빅데이터를 기반으로 청소년 자살예방대응체제를 구축한 사례를 볼 수 있다. 대표적으로 미국 뉴욕주 정신보건국은 자살통계 및 데이터에 근거하여 자살예방 활동을 하기 위해 “정신보건서비스 및 임상지식 향상시스템(PSYCKES: Psychiatric Services and Clinical Knowledge Enhancement System)”을 활용하고 있다.⁵⁸⁾

PSYCKES는 메디케이드(Medicaid) 정보를 활용한 웹 기반 데이터베이스로서 정신건강 또는 약물 남용 서비스를 이용한 의료보험 사용자와 정신과 처방 이력을 가진 8백만명 이상의 개인 의료정보를 축적하고 있으며, 정보 접근권한이 있는 사용자가 5년간의 기록을 확인할 수 있는 실시간 통합 의료기록 허브로 활용되고 있다. PSYCKES에서 검색할 수 있는 개인의료정보는 연방정부의 개인건강정보보호법(HIPAA)에 따라 철저히 보호 받지만 개인의 자살 시도, 자살 사고, 자해·음독에 관한 정보 및 자살 선별 도구와 안전계획 수립 유무 등을 확인할 수 있으며 현재 뉴욕주 임상 현장에서 내담자의 자살 예방에 매우 중요하고 실질적인 도구로 활용되고 있다(손해인, 2019).

빅데이터는 또한 청소년비행 등의 범죄예측 시스템에도 활용되고 있다. 미국의 범죄예방시스템은 프레드폴(PredPol), IBM 통계분석 범죄예측예방시스템(IBM SPSS Crime Prediction and Prevention), 헨치랩(HunchLab) 등 여러 가지가 있지만, 그 중에서 특히, PredPol과 Hunch Lab은 청소년범죄예방과 관련한 빅데이터 활용 사례로 제시될 수 있다. 특히, 시카고 7구역에서 활용한 Hunch Lab의 사례는 아동 청소년만을 대상으로 한 것은 아니지만, 청소년을 포함하여 지역사회 범죄율을 낮추는데 빅데이터가 기여한 것으로 평가된다(Bachner, 2013).

58) https://omh.ny.gov/omhweb/psyckes_medicaid/about/ 에서 2020년 8월 3일 검색

실제로 PredPol은 미국의 로스엔젤레스, 애틀랜타, 산타크루즈 등 여러 도시에 서 적용하고 있는데, 이들 지역의 범죄 예방률이 기존 대비 15-20% 정도 향상된 것으로 나타났다(Thompson, 2011). 또한, HunchLab은 지역 내 사업장 위치, 날씨 정보, 경제 정보 등의 다양한 요인을 분석하고 범죄를 예측하는 시스템으로 필요시에는 지역 내에 설치된 충격음 감지 시스템도 동원하며, 마약 거래와 갱단 구성원 등과 관련한 데이터들을 활용할 수 있다. 이를 분석해 범죄 발생 우려가 높은 곳에는 적절히 순찰차를 배치해 범죄 발생률을 줄인 것이다. HunchLab은 St. Louis County, 아리조나 주 Peoria, 네브라스카 주, Lincoln, 델라웨어 주 New Castle County 등에서 사용 중이며, 필라델피아, 샌프란시스코, 덴버, 워싱턴 등의 대도시 일부에도 비슷한 시스템을 도입해 운영 중이다(Leiter, 2013).

4) 우리나라의 정책적 시사점

다양성에 기초를 둔 미국사회의 특징은 청소년에 대한 미국의 국가적 관심에서 도 다양성을 인정하는 가운데 개인적 성장과 미국사회의 공동목표 추구를 추구하는 것으로 나타난다. 아동·청소년관련 데이터를 취합하는 일에서도 미국은 국가전략의 큰 틀에서 자율적이고 다양한 형태를 보이고 있다. Data.gov로 대표되는 공공서비스를 제공하고 있는 오픈데이터(open data) 정책을 실시하고 있으며, childstats.gov, youth.gov 및 Child Trends와 KIDS COUNT 등의 다양한 빅데이터를 가지고 있다. 이제 지금까지 논의한 내용을 바탕으로 우리나라의 정책 개발을 위한 시사점을 다음과 같이 몇가지로 구분하여 제시하기로 한다.

첫째, 데이터 생산(production)과 관련하여 양질의 데이터를 체계적으로 구축할 수 있는 제도 마련과 함께 지속적인 자료축적이 가능하도록 하는 법적 장치를 확보할 필요가 있다. 빅데이터의 성공적 활용을 위해서는 수준높은 다양한 데이터를 축적하는 것이 중요하며, 각종 관련 기관별, 분야별, 영역별로 각기 다른 항목,

형태, 및 구조로 형성되는 자료들을 보다 실제적이고 유용한 데이터가 되도록 해야 한다. 자료를 구축함에 있어서 주제별, 내용별로 체계적인 지표를 통해 일관성있게 지속적으로 생산하기 위해서는 청소년기본법 등의 관련법을 통해 구체적인 내용을 명시할 필요가 있다. 실제로 우리나라의 경우, 청소년시설의 운영에서 빅데이터를 구축하고 활용할 필요성이 제기되고, 청소년단체와 시설에서는 프로그램에 참여한 청소년들과 수요자들에 대한 요구와 반응을 데이터베이스화할 것도 강조되고 있다(이경상, 이창호, 김민, 2019). 그러나 현실적으로 청소년시설과 단체 수준에서 관련 프로그램과 참여자 등에 관한 정보를 자료화하기는 쉽지 않으며, 이를 위한 기반 프로그램을 만들어 제공하거나 제공자료의 항목, 내용, 형태와 구조 등에 대한 공통적 체계성을 찾거나 데이터 제출을 의무화하는 규정은 없는 실정이다.

미국의 경우, 공공부문의 각종 데이터들이 다양한 주제와 형태로 생산되고 있으며 데이터 공개를 의무화하는 법안이 마련됨으로써, 민간부분의 경우에도 여러 가지 형태로 일관적인 항목과 수집방법을 통해 지속적으로 생산되고 있음을 알 수 있다. 이는 2009년 5월 출범했던 Data.gov의 경우 최초 47개의 데이터베이스로 출발했지만, 11년이 지난 2020년 8월에는 230,000개가 넘는 데이터셋을 구축한 것으로도 알 수 있으며 이러한 발전에는 “OPEN 정부데이터법(OPEN Government Data Act)”이 크게 기여하였다고 본다. 아동·청소년건강관련 공공 데이터로 볼 수 있는 healthdata.gov에서 취합되고 있는 데이터셋의 경우에도 약물남용(substance abuse)과 우울(depression) 등 50여개의 연관주제어를 활용하여 데이터를 검색할 수 있으며, 데이터 포맷도 html, pdf 등 14개 이상의 형태로 제공되고 있음을 알 수 있다. 공공과 민간부문 모두 데이터의 성격은 매우 구체적이고 실제적인 내용을 담고 있는 상황이다. 이런 점에서 우리나라의 경우, 기존의 각종 현황자료나 기초자료 및 연구자료들이 보다 구체적인 내용을 담고 지속적으로 확보될 수 있도록 형태와 항목 등의 기준을 마련하여 체계적 생산을

담보할 수 있는 제도화가 요망된다.

둘째, 데이터의 제공(provision)과 관련하여 생산된 데이터를 종합적으로 관리하고 개방적으로 활용할 수 있는 데이터 플랫폼 구축과 효율적 관리·운영 체제를 마련할 필요가 있다. 미국 연방정부는 data.gov 또는 healthdata.gov라는 포털 플랫폼을 구축하고 공공데이터를 제공하면서 정보의 접근성과 활용도를 촉진하기 위하여 많은 노력을 기울이고 있다. 이는 정부 주도의 빅데이터 기반 공공정책을 추진하는 것과 함께, 각종 빅데이터 장비와 소프트웨어들이 주로 미국을 중심으로 각종 기관과 기업 등을 통해 개발·제공되고 있는 것과도 관련된다(이지혜, 제미경, 조명지, 손현석, 2014).

우리나라에도 공공데이터는 data.go.kr 포털을 운영하고 있으나, 2020년 8월 현재 이곳의 아동·청소년 데이터는 각종 실태연구나 일반현황 관련 자료들을 중심으로 제공되고 있으며, 청소년백서에서 제시되고 있는 내용을 넘어서는 보다 실제적·구체적 자료들은 찾아보기는 쉽지 않다. 연구자료와 민간기관에서의 자료가 체계적이며 효율적으로 관리되고 있는 경우도 많지 않으며, 관련 데이터를 활용하고자 할 경우에도 접근이 제한되고 승인을 받는 등의 절차적 과정이 선행되기도 한다. 이런 점에서 미국의 “OPEN 정부데이터법”에서 제시하고 있는 “개방적(open), 공공적(public), 전자적(electronic), 필요한(necessary)” 의미를 구현할 수 있는 데이터 제공을 위한 청소년관련 정보를 집대성하는 종합 공공데이터 포털을 개발하는 등의 정책개발이 있어야 할 것이다.

셋째, 데이터의 생산과 제공에 관한 이슈 다음으로는 데이터 활용도 제고(promotion)를 위한 연계·공유적 연동 시스템을 구축하는 일이 요망된다. 우리나라에서는 현재 각종 청소년관련 기관을 중심으로 “공공데이터의 제공 및 이용 활성화에 관한 법률”에 따라 데이터 공개와 활용을 위한 노력을 기울이고 있으나 각 기관별로 개별·분산적이며 기관간 상호 연계적 운영은 결여하고 있다. 대표적인 기관으로 한국청소년정책연구원의 경우에는 패널조사 데이터를 비롯하여 횡단

연구 자료를 공개하고 청소년통계아카이브센터를 구축하고 있으며, 한국청소년단체협의회, 청소년활동진흥원, 한국청소년상담복지개발원 등의 경우에도 청소년단체/시설 DB, 빅데이터센터, 빅데이터분석 등의 이름으로 관련 자료들을 제공하고 있음을 확인할 수 있다. 그러나, 이들 기관의 데이터 생산과 제공의 경우도 미국의 사례에서 보는 것과 같은 하나의 기관간 협력체제(예: the Forum)가 주관하거나 또는 데이터 온톨로지나 합의된 지표체계에 따라 지속적으로 공포하는 시스템(예: youth.gov: Child Trend: KIDS COUNT)에 따라 수행되고 있는 것은 찾아보기 어렵다.

더욱이 아동·청소년과 관련한 보건·의료 분야 등의 데이터의 접근이나 정보공유를 위한 노력은 여전히 제한적이다. 이러한 데이터는 개인정보가 중요하고 정보활용에 한계가 있지만, 그에 필요한 제도적 장치를 마련하고 관련 데이터를 공유할 수 있는 방안을 찾아보는 노력이 있어야 할 것이다. 특히 강조할 점은 미국의 경우에서 볼 수 있는 청소년문제 위험예측 대응체제 구축과 관련한 시스템 공유 등과 같은 사례를 우리의 경우는 찾아보기 어렵다는 것이다. 미국의 PSYCKES의 경우, 그것이 빅데이터로서 갖는 의미는 메디케이드 보험청구 시스템을 비롯하여 개인 의료정보를 검색할 수 있는 다양한 전산망과 연동하면서 자살시도자 및 자살사고자를 관리할 수 있다는 점이다. 우리나라의 경우도 개인정보보호법에 의해 정보공유의 어려움은 있지만, 현행의 정신건강사례관리 시스템과 국민건강보험공단 등의 정보망과 국가응급의료전산망 등이 정보를 공유할 수 있는 연동시스템을 갖출 수 있다면, 자살시도자 및 자살 사고자를 파악하고 예방관리할 수 있는 방법을 찾을 수 있을 것으로 본다(손해인, 2018).

미국의 범죄 예방 시스템 활용사례와 같은 것은 우리나라에서도 유사하게 활용하고 있지만, 이러한 시스템들을 청소년문제나 비행위험 예측을 위해 활용할 수 있도록 함으로써 그 기능과 효용성을 개선할 필요가 있다. 데이터 활용도 촉진을 위한 공유시스템 구축은 정부기관 뿐만 아니라, 학교, 아동·청소년단체와 시설

및 연구개발기관 등의 각종 성과물들을 체계적으로 관리하고 그 활용도를 높일 수 있게함으로써 관련 연구를 촉진하는데 기여할 것이다. 이런 점에서 향후, ① 우선 기관별로 분산적으로 제공하고 있는 자료들을 하나의 플랫폼을 통해 종합제공하여 필요로 하는 경우 누구나 하나의 포털에서 보다 쉽게 접근하고 유용하게 활용할 수 있는 공유시스템을 마련하고 ② 미국에서와 같은 주요 기관간 협력체제(the Forum)를 구축하여 공신력 있는 시스템 운영 방안을 마련한 후 ③ 점차 청소년 빅데이터 전담기관을 설립하는 등의 단계적이며 실질적인 방법의 정책적 접근이 있어야 할 것이다.

넷째, 빅데이터 활용을 높이기 위한 노력은 그에 따른 개인정보의 보호(Protection)와 관련한 법적 장치를 마련하는 일과 불가분의 관계가 있다. 특히, 빅데이터 중에서도 가장 많은 부분을 차지하는 비정형 데이터의 경우 스마트폰이나 소셜 미디어를 기반으로 한 다양한 개인 정보를 포함하고 있다는 점에서 해당 정보의 수집과 활용에 앞서 관련 정보의 제공과 관련한 권한과 책임의 문제, 정보 유출과 보안의 문제 등에 대한 법·제도적 보호장치가 선행될 필요가 있다. 실제로 빅데이터 기술이 발전함에 따라 청소년문제의 예방, 발견 및 개입을 위한 기회가 더욱 증대되고 있지만, 그와 동시에 프라이버시와 관련한 함의들이 무시할 수 없는 중요한 이슈로 대두되고 있다.

따라서 빅데이터 전략을 실행할 때는 시민적 자유(Civil Liberties)를 위반할 잠재성이 따르는 것이 가장 큰 관심사가 되고 있다는 점에 유의할 필요가 있다. 이는 특히 미국의 경우 등록되지 않은 이민자나 동성애자 등과 같은 취약한 구성원에 속하는 청소년들에게는 더욱 중요한 문제가 되고 있다. 예를 들면, 사회적 정서적 스트레스 요인을 가진 취약 청소년들은 학교 당국이 자신들을 빅데이터를 통해 모니터링하고 있다는 것을 알게 되면 학교를 떠나게 되는 경우도 나타나고 있다(Watson & Christensen, 2017). 이에 따라, 개인정보보호와 관련하여 미국 연방정부 교육부는 디지털 시대 학생들의 프라이버시(사생활보호)를 보호하기 위한

필수요건(requiremen)과 최상의 실천(best practices) 기준을 마련하였다. 학생들에게 투명하게 공개하고(transparent), 부모 동의가 적절하다는 점을 고려하고, 학생정보를 보호하기 위해 제정된 FERPA 같은 법에 대한 자각을 유지하는 노력을 하고 있다. 그러나, 이러한 노력에도 불구하고 여전히 상당한 혼란들이 남아 있으며, 새로운 학생 사생활 보호법이 필요하다고 주장하고 있는 실정이다.

다섯째, 데이터의 생산·제공·활용 및 개인정보 보호 등에 대한 이해 증진은 물론, 빅데이터와 관련한 우리나라의 국가미래전략을 고려한 청소년분야의 준비(preparation)가 필요하다. 국내에서도 빅데이터 기술 인력을 양성하기 위한 각종 교육 인프라와 실습 기회 등을 다양하게 제공하고 있지만, 청소년관련 기관 종사자와 청소년단체·시설의 청소년지도자 등에 대한 데이터 생산·가공·제공·활용 및 정보보호 등과 관련한 역량함양을 위한 교육기회는 매우 부족한 현실이다.

미국의 경우, 정부주도의 추진전략으로 마련된 “빅데이터 R&D 전략계획”에서 빅데이터 인력을 7대 전략 중 하나로 제시하고 있으며, 4대 빅데이터 지역혁신 허브 중에서 북동부 허브를 중심으로 교육관련 데이터사이언스를 구축하고 있으며, 북동부 학생데이터단(Student Data Corps)를 만들어 전 지역 학생들에게 교육하고 있다. 우리나라에서도 다양한 형태로 유사한 교육이 대학 등에서 실시되고 있으나, 아동·청소년과 관련한 분야에서 이러한 프로그램이 제공되고 있는 것을 찾기는 쉽지 않다. 따라서, 청소년지도자 연수과정을 통해 빅데이터에 대해 공부하고, 청소년단체와 청소년시설 등의 종사자를 대상으로 하는 직무교육과 보수교육 등을 통해서 데이터 생산·제공·활용 및 정보보호의 중요성에 대한 지식과 구체적인 실무역량을 함양하고 전문직 윤리에 대한 인식을 강화하는 방안을 강구할 필요가 있다.

미국국토안보부에서 조사한 2019년부터 2022년까지의 분석결과를 보면, 빅데이터 시장은 매년 17.5%씩 성장하며 국가안보와 공공 안전을 위해 가장 많이 사용될 것으로 전망하고 있으며, 빅데이터를 가장 많이 사용할 것으로 예상되는

상위 국가 중에 미국은 1위, 한국은 11위로 제시되고 있다(USA Homeland Security, 2018; 한국정보화진흥원, 2019, 재인용). 아동과 청소년과 관련한 데이터는 4차 산업혁명을 바라보는 현 상황에 비추어 정형 데이터 뿐만 아니라, 일상에서 만들어지는 비정형 데이터까지 포함할 때 그 중요성은 다시 말할 필요가 없을 것이다. 이를 위해 우리나라 아동·청소년 데이터의 생산(production)과 제공(provision), 활용도 제고(promotion), 정보보호(protection) 및 그에 따른 아동·청소년 분야의 준비(preparation)는 시급하면서도 중차대한 국가발전 의제(agenda)로 논의되어야 할 것이다.

특히 빅데이터 분야는 미국을 비롯한 선진국들을 중심으로 아직은 초기 단계에서 서로 경쟁하고 있는 상황으로 볼 수 있다는 점에서 빅데이터 활용과 관련한 이슈와 기술개발의 문제는 우리나라의 대표적인 국가적 과제이며, 이 분야에 대한 R&D 투자는 매우 긴요한 과업이 되고 있다. 아동·청소년 분야는 국가발전의 주요한 인력자원과 관련한 정책 영역이라는 점에서 향후 이와 관련한 빅데이터 체계에 대한 연구와 정책개발 및 보다 적극적인 활용 방안을 마련하는 일이 그에 따른 개인정보보호에 대한 관련 정책검토와 함께 더욱 다양하게 수행되기를 기대한다.

2. 영국의 아동·청소년 빅데이터 체계 구축 및 활용 사례

1) 데이터 구축 및 공개에 관한 주요 정책

영국은 공공 정보를 공개하는 오픈 데이터 정책(open data policy)을 선도하고 있는 대표적 국가로, 2009년 「스마트 정부 구현을 위한 실행계획서(Putting the Frontline First: Smarter Government)」 발표를 시작으로 2012년 「오픈 데이터 전략(Open Data Strategy)」, 2013년 「데이터 역량 강화 전략(Seizing the Data Opportunity: A Strategy for UK Data Capability)」, 2014년 「오픈 데이터 전략 2014-2016(Open Data Strategy 2014-2016)」 등을 발표하였다. 2017년에는 세계를 이끄는 디지털 경제 구축을 목표로 하여 「영국 디지털 전략(UK Digital Strategy)」을 발표하는 한편 디지털 기술에 기반하여 정부 혁신 및 정부와 국민의 관계 개선을 추구하며 「정부 혁신 전략(Government Transformation Strategy)」을 발표하였고, 현재는 세계를 이끄는 데이터 기반 경제국을 목표로 「국가 데이터 전략(National Data Strategy)」을 설계 중이다. 아래 표로 그간의 데이터 구축과 공개에 관하여 영국 정부가 추진해 온 정책들을 요약하였다.

표 IV-3 영국의 데이터 구축 및 공개 관련 주요 정책

시기(연도)	정책
2009	스마트 정부 구현을 위한 실행계획서(Putting the Frontline First: Smarter Government) 발표
2010	공공 데이터 포털이라 할 수 있는 'data.gov.uk' 개설
2012	오픈 데이터 백서(Open Data White Paper) 및 오픈 데이터 전략(Open Data Strategy) 발표

시기(연도)	정책
2013	데이터 역량 강화 전략(Seizing the Data Opportunity: A Strategy for UK Data Capability) 발표
2014	오픈 데이터 전략 2014-2016(Open Data Strategy 2014-2016) 발표
2017	영국 디지털 전략(UK Digital Strategy) 발표 영국 정부 혁신 전략(Government Transformation Strategy) 발표
2018	data.gov.uk 및 Find Open Data 서비스 론칭
2019 - 2020	국가 데이터 전략(National Data Strategy) 설계 중 (2019년 현황조사 및 원탁회의 시행, 2020년 발표 예정)

* 출처: 영국 정부에서 발표한 각 보고서 내용을 종합하여 연도별로 정리

① Open Data Strategy (오픈 데이터 전략, 2012 / 2014-2016)

영국은 공공기관의 기록정보에 대한 국민의 알 권리와 접근권을 보장하는 2000년 정보공개법(Freedom of Information Act 2000) 제정을 기점으로 공공 데이터 개방 및 그를 통한 사회 혁신을 단계적으로 추진하였다. 첫 번째로 2009년 12월 「스마트 정부 구현을 위한 실행계획서(Putting the Frontline First: Smarter Government)」를 발표하며 데이터(특히 공공 데이터) 개방을 통해 정부 운영의 투명성과 효율성을 향상하고 사회 혁신을 추구하겠다는 의지를 표명하였고(영국 정부포털), 그 결과로 2010년 누구나 영국 정부의 데이터에 접근할 수 있는 'data.gov.uk'를 개설하였다(영국 오픈 데이터 포털). 2012년에는 「오픈 데이터 백서(Open Data White Paper)」와 「오픈 데이터 전략(Open Data Strategy)」을 발표하였는데 이를 통해 공공 데이터 개방에 관한 정부의 방침과 실적, 계획을 소개하였고 실제로 의료, 고용, 세금, 교육 등의 공공 데이터들을 순차적으로 개방하였다(Department for Business, Innovation & Skills, 2014).

이후 영국 정부는 2013년 영국을 포함한 G8 국가들이 합의, 발표한 「G8 오픈 데이터 헌장(G8 Open Data Charter)」의 내용과 원칙에 따라, 기존의 오픈 데이터 전략을 수정하며 2014년 새로운 오픈 데이터 전략(「Open Data Strategy 2014-2016」)을 발표하였다. 새로운 오픈 데이터 전략은 (i) 데이터 공개의 기본 원칙, (ii) 데이터의 양과 질, (iii) 포괄적 이용가능성, (iv) 거버넌스 개선을 위한 데이터 공개, (v) 혁신을 위한 데이터 공개의 5대 원칙으로 이루어져 있다 (Department for Business, Innovation & Skills, 2014).

■ 원칙 1 - 데이터 공개의 기본 원칙(Open Data by Default): 정부와 공공기관 등의 데이터는 가능한 한 공개함을 기본 원칙으로 한다. 기관 및 데이터의 특성상 데이터를 완전히 공개하기 어렵거나 일부 공개조차 불가능할 수 있는데, 이러한 예외적 상황은 각 사례별로 접근하여 공개 여부를 결정한다. 또한 예외적 상황의 경우에도 사회에 긍정적으로 기여할 수 있는 연구 목적의 경우에는 연구자들이 최대한 접근할 수 있도록 해야 하며, 다만 어떠한 경우에도 개인 정보를 비롯하여 민감한 정보를 포함한 데이터는 철저히 보호되어야 한다.

■ 원칙 2 - 데이터의 양과 질(Quality and Quantity): 정부 및 공공기관을 통해 공개하는 데이터는 '확장된 정부 전략(Wider Government Strategy)'의 요구사항, 즉 보다 많은 데이터를 공개하는 동시에 공개하는 데이터의 질(quality)이 높고, 데이터 공개를 통해 힘과 권력을 소수에서 대중으로 이동시키는데 기여해야 한다. 또한 공개하는 데이터는 통계적으로 강건하고(robustness) 타당하며(validity), 쉽게 이용할 수 있고(availability) 투명하면서(transparency) 개인 정보를 철저히 보호해야 한다.

■ 원칙 3 - 포괄적 이용가능성(Usable by All): 정부 및 공공기관을 통해 공개하는 정보와 데이터는 모두가 접근할 수 있고 이용할 수 있어야 한다. 이에 정부는

홈페이지, SNS, 유튜브 등 여러 디지털 채널을 활용하여 보다 많은 사람들이 데이터에 접근할 수 있도록 하고 있으며, 동시에 데이터가 보다 이용하기 용이한 형태로 공개될 수 있도록 하는데 주의를 기울이고 있다.

■ 원칙 4 - 거버넌스 개선을 위한 데이터 공개(Releasing Data for Improved Governance): 정부 및 공공기관의 데이터 공개를 통해 국민들은 해당 부처, 기관의 성과와 생산성, 효율성에 대해 보다 잘 이해할 수 있고, 자신들이 납부한 세금이 어디에 어떻게 사용되었는지 파악할 수 있다. 예를 들어 기업혁신기술부(Department for Business, Innovation & Skills, BIS)의 공개 데이터 세트를 보면, 재정지출, 조직도 및 직원현황, 계약업체, 에너지 이용, 고위급 공무원의 법인카드 및 출장 경비 내역 등이 매월 또는 분기별로 공개되고 있다. 이러한 데이터 공개를 통해 정부의 투명성과 책무성을 제고하며 거버넌스를 개선하는 것이 네 번째 원칙이다.

표 IV-4 기업혁신기술부(BIS)의 공개 데이터(예시)

BIS 데이터 명	공개 시기
재정 지출 데이터(모든 지출 항목)	매 월
정부 조달 카드 지출	매 분기
ICT, 홍보마케팅, 자문, 채용, 자산 등의 지출	매 분기
조직도 및 직원 현황	연 2회
사무소 관리 정보	매 월
장관의 지출 정보(경비, 출장, 접대 등)	매 분기
고위 임원의 지출 정보(경비, 출장, 접대 등)	매 분기
자문 및 미디어 등에 대한 접대와 회의 지출	매 분기
계약업체 정보	상시
실시간 에너지 사용	매일
분기별 데이터 요약	매 분기

* 출처: Department for Business, Innovation & Skills (2014). Open Data Strategy 2014-2016: 13.

■ 원칙 5 - 혁신을 위한 데이터 공개(Releasing Data for Innovation): 마지막 원칙은 데이터 공개가 영국 경제와 사회에 미치는 영향에 관해서다. 영국 정부는 국가와 사회 혁신을 위한 데이터 공개를 위해 공공과 민간 부문, 학계 등의 데이터 역량 강화, 다양한 주체들과의 협력관계 개발, 데이터를 활용한 새로운 기회 모색 및 그를 위한 리더십 준비 등의 노력을 기울이고 있다.

한편 영국은 이러한 데이터 공개를 위하여 2010년부터 'data.gov.uk' 사이트를 운영하여 사람들이 쉽게 정부 데이터를 검색하고 이용할 수 있도록 하고 있으며, 2018년 3월에는 'data.gov.uk' 사이트를 리노베이션하며 'Find open data'라는 오픈 데이터 검색 서비스를 출시하였다(영국 오픈 데이터 포털). 'Find open data' 서비스에서 사람들은 (i) 중앙정부, 지방정부 및 공공기관 데이터를 직접 검색하고 (ii) 제공되는 링크를 통해 파일을 다운로드할 수 있으며 (iii) 데이터를 직접 발행할 수 있는 계정을 생성하여 데이터를 발행, 공유할 수도 있다. 'data.gov.uk' 플랫폼 및 'Find open data' 서비스는 정부디지털서비스청(Government Digital Service, GDS), 국가통계사무소(Office for National Statistics, ONS), 디지털·문화·미디어·스포츠부(Department for Digital, Culture, Media & Sport, DCMS) 그리고 국가사이버보안센터(National Cyber Security Centre, NCSC) 등이 협력하여 운영하며, 이 중 GDS가 주 담당 및 조정 역할을 맡고 있다. 아래 그림으로 오픈 데이터 포털인 'data.gov.uk'의 'Find open data' 서비스를 소개하였는데 경제 및 산업, 정부, 재정지출, 안보, 도시와 지역사회, 사회, 교통, 범죄와 사법, 건강, 교육, 환경, 지도 등이 주요 범주임을 확인할 수 있다.

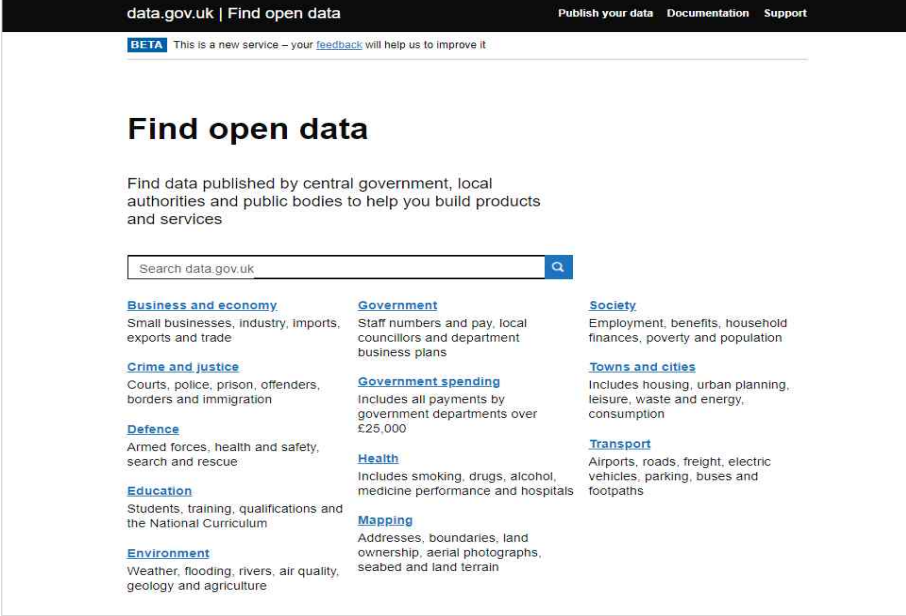


그림 IV-1 영국 오픈 데이터 포털 'data.gov.uk'

② A Strategy for UK Data Capability(영국 데이터 역량 강화 전략, 2013)
 빅데이터를 새로운 가치 창출의 기회로 주목하며 오픈 데이터 정책을 일찍부터 시도해온 영국은 2013년 10월 「데이터 역량 강화 전략(Seizing the Data Opportunity: A Strategy for UK Data Capability)」을 발표하였다. 본 전략은 데이터 구축과 분석, 해석, 관리, 보안 등과 관련한 역량 강화에 주안점을 둔 정책으로, 정보경제위원회(Information Economy Council, IEC)와 E-인프라 리더십위원회(E - infrastructure Leadership Council)가 공동으로 주관하였다(Department for Business, Innovation & Skills, 2013). 본 보고서 제목 'Seizing the Data Opportunity(데이터 기회 포착)'에서 추측할 수 있듯이, 영국은 국가적 차원에서 데이터 역량을 강화하여 글로벌 데이터 강국으로서의 성장동력을 얻고자 하였는데, 이때 데이터 역량(data capability)을 (i) 인적 자원, (ii)

데이터 저장과 분석을 가능하게 하는 도구와 인프라, (iii) 가능성과 기회를 가져오는 데이터 그 자체의 세 가지 차원으로 이해하였다(Department for Business, Innovation & Skills, 2013).

■ 인적 자원(Human Capital): 인적 자원 관점에서의 데이터 역량은 데이터 기술을 갖춘 노동자이자 데이터에 자신감을 갖춘 시민(A skilled workforce, and data-confident citizens)으로 이해할 수 있다. 데이터 기술(Data skills)은 데이터를 구축하고 분석, 해석, 소통, 보호, 관리 등을 할 수 있는 기술이라 할 수 있고, 데이터 자신감이라 함은 데이터를 검색하고 분석하고 해석할 수 있으며 데이터와 관련한 자신의 권리를 인식하고 발휘할 수 있는, 다시 말해 데이터 리터러시(data literacy)로 이해할 수 있다.

■ 도구와 인프라(Tools and Infrastructure): 도구와 인프라 관점에서 데이터 역량은 데이터를 효과적으로 생성하고 저장, 분석, 관리할 수 있는 인프라, 소프트웨어, 민-관-학-연의 R&D 협력체계 등으로 설명할 수 있다.

■ 가능성과 기회를 가져오는 데이터 그 자체(Data itself as an enabler): 데이터 역량 강화의 또 다른 핵심 요소는 데이터 자체의 접근성과 안정성에 있다. 즉 데이터를 이용하려는 사람이 누구든지 데이터에 자유롭게 접근할 수 있고 안정적으로 활용할 수 있어야 한다.

■ 이러한 이해에 기초하여 영국 정부는 데이터 전문 인력 양성, 데이터 인프라와 도구 확보, 데이터 협력체계 강화, 데이터의 접근성과 안정성 제고 등을 위한 여러 실행계획을 추진하고 있는데, 그 가운데 데이터를 둘러싼 민-관-학-연 주체들 간의 협력체계 관련 계획이 상당하다는 점은 주목할 만하다. 예를 들어 정부는 데이터 분석 기술과 역량에 대한 교육시스템을 확립하여 데이터 전문 인력을 양성하고자 기업, e-Skills UK, 대학들, 오픈 데이터 연구소(Open Data Institute, ODI) 등과 협력체계를 구축, 강화하고 있고 데이터 인프라 강화를 위해 Data Centre Alliance, Intellect UKTI, E-Infrastructure Leadership Council 등

과의 협력을 추진하고 있으며, 연구 목적의 데이터 접근과 활용을 지원하고자 교육 및 연구기관을 비롯한 다양한 연구 커뮤니티와 협력하고 있다(Department for Business, Innovation & Skills, 2013: 47).

③ UK Digital Strategy(영국 디지털 전략, 2017)

2017년 Department for Digital, Culture, Media & Sport(DCMS, 디지털·문화·미디어·스포츠부)는 「UK Digital Strategy(영국 디지털 전략)」을 발표하였다. ‘UK Digital Strategy’는 세계를 이끄는 디지털 경제를 구축하는 것을 목표로 하여 (i) 영국 전역에 세계 수준의 디지털 인프라 구축(Connectivity - Building world-class digital infrastructure for the UK), (ii) 모든 사람에게 필요한 기술에 대하여 접근할 수 있는 기회 제공(Digital skills - Giving everyone access to the digital skills they need), (iii) 디지털 비즈니스의 창업과 운영을 위한 최적의 환경 조성(Digital sectors - Making the UK the best place to start and grow a digital business), (iv) 모든 비즈니스의 디지털 비즈니스로의 확장 지원(Wider economy - Helping every British business become a digital business), (v) 안전하게 일하고 생활할 수 있는 디지털 환경 조성(A safe and secure cyberspace - Making the UK the safest place in the world to live and work online), (vi) 세계 최고의 전자정부 운영(Digital government - Maintaining the UK government as a world leader in serving its citizens online), (vii) 데이터 공개 확대 및 대중의 신뢰 향상(Data - Unlocking the power of data in the UK economy and improving public confidence in its use)의 7가지 전략을 수립, 추진하고 있다(Department for Digital, Culture, Media & Sport, 2017).

이 중 데이터에 관한 전략을 구체적으로 살펴보면 영국 정부는 데이터 공개 확대 및 역량 강화를 위하여 데이터 인프라 구축, 공공 및 민간 부문의 데이터

기술 및 역량 강화 지원, 국가 간 데이터 공유와 교류 등의 사업들을 추진하는 한편, 데이터 이용과 관리에 관한 대중의 신뢰를 확보하고자 개인정보감독기구(Information Commissioner's Office, ICO)를 통해 개인의 정보기본권을 보호하고 정부와 기업 등의 개인정보 보호와 관리 책임을 강화하고 있으며, 데이터의 윤리적 활용을 위한 구체적인 방침과 체계 마련도 추진하고 있다(Department for Digital, Culture, Media & Sport, 2017).

④ Government Transformation Strategy(정부 혁신 전략, 2017)

Government Digital Service(GDS, 정부디지털서비스청)는 디지털 기술에 기반하여 정부와 국민의 관계를 혁신하고, 국민의 욕구와 필요에 보다 부합하는 정부가 되고자 2017년 「Government Transformation Strategy(정부 혁신 전략)」을 발표하여 추진하고 있다. 본 전략은 정부와 국민의 관계 혁신을 위해 전자 정부 서비스를 비롯해 다양한 방안들을 제시하는데, 그 중 하나가 '데이터 이용 개선(better use of data)'으로 공공 데이터의 공개 확대, 정부 내 효과적 데이터 이용 증진을 위한 장벽 개선, 데이터 이용에 관한 제반 사항을 이끌어 나갈 '최고데이터관리자(Chief Data Officer)' 임명, '데이터자문위원회(Data Advisory Board)' 신설, 데이터의 전문적·효과적 이용을 위한 데이터과학 및 분석 역량 제고, 데이터 이용 및 공유에 관한 윤리의식 함양, 국가 차원의 데이터 인프라 구축 등의 과제들을 포함한다. 아래 표로 데이터 이용 개선과 관련하여 중점 과제로 제시된 내용을 제시하였다.

표 IV-5 데이터 이용 개선을 위한 중점 과제(2017 - 2020)

- 공공 데이터 공개와 공유 및 효과적인 데이터 활용을 위한 장벽 제거
 - 2017년 디지털경제법 제정 및 시행
- 정부 내 데이터 과학 영역 확장과 구성원들의 데이터 역량 강화를 통해, 의사결정 과정 내 데이터 활용 증진
- 데이터 기반의 전문적 정책 분석을 담당하는 팀 구성(데이터 분석가, 행동과학자 등), 정부 통계 및 그 외 다양한 데이터 생성 방식 개선, 정부 통계 및 데이터의 빠르고 적절한 공개와 공유, 정부 내 데이터 과학 및 분석 역량 강화
- 안전하고 적절한, 그리고 윤리적인 데이터 관리와 이용
- '최고데이터관리자(Chief Data Officer)' 임명, '데이터자문위원회(Data Advisory Board)' 구성 및 운영, 안전하고 윤리적인 데이터 관리에 관한 지침 마련과 교육, 데이터 관련 시민의식 함양 등
- 국가 차원의 데이터 인프라 구축 - 모든 공공기관의 기록정보가 안전하게 저장, 관리되는 레지스터
- 정부 및 공공기관의 데이터 공개
- 정부 서비스 내 API(애플리케이션 프로그래밍 인터페이스) 적용 확대
- 데이터 검색 및 발견을 위한 도구 개선
- 정부의 주요 데이터 저장소 내 데이터가 저장, 관리되는 방식의 혁신

* 출처: Government Digital Service (2017). Government Transformation Strategy: 48-49.

⑤ National Data Strategy(국가 데이터 전략, 2020년 발표 예정)

2018년 6월 Department for Digital, Culture, Media & Sport(DCMS, 디지털·문화·미디어·스포츠부)는 국가 차원의 강력한 데이터 생태계를 구축하고 모든 기관(조직)과 시민의 데이터 역량 및 접근성을 제고하고자 2020년까지 「National Data Strategy(NDS, 국가 데이터 전략)」을 수립하겠다고 발표하였다(영국 정부포털). 이를 위하여 2019년 현황조사(Call for Evidence)를 실시하였는데, 학교와 연구기관, 시민사회, 기업과 소상공인, 스타트업, 공공 부문 및 시민(개인) 등 다양한 이해관계자를 모두 포함하였다.

이후 DCMS는 현황조사 결과를 바탕으로 다양한 이해관계자들이 참여하는 원탁회의를 20회 이상 진행하였는데, 학계를 비롯하여 기업, 공공부문 및 시민사회 등 250여 곳에서 참여하였다(영국 정부포털). 원탁회의 주제는 크게 (i) 모든 개인

과 조직이 데이터 기반 경제와 사회에서 효과적으로 참여할 수 있도록 보장하는 것, (ii) 효과적인 데이터 수집과 공유, 이용을 통해 성장과 생산성, 공공서비스 개선을 도모하는 것, (iii) 공공 및 민간 부문에서 데이터 공유와 이용에 관한 협력과 조정을 향상함으로써 효율성을 제고하는 것, (iv) 개인과 조직이 믿을 수 있는 방법으로 데이터가 이용되도록 보장하는 것의 4가지였으며, 이들 주제에 대하여 참여자들은 활발히 토론에 참여하였다(영국 정부포털).

아래 표와 그림으로 국가 데이터 전략 수립을 위한 현황조사 질문 및 원탁회의 활동모습을 제시하였다.

표 IV-6 국가 데이터 전략 수립을 위한 현황조사 질문

대영역	소영역	주제	질문
사람	신뢰	기회와 장벽	1-1. (공공, 민간, 비영리) 기관은 데이터 이용의 신뢰성을 어떻게 입증할 수 있는가? 1-2. 기관에서 제공하는 정보가 어떻게 이용되는지 일반 시민이 쉽게 알 수 있는가? 1-3. 기관은 개인정보를 이용함에 있어 신뢰를 잃을 수 있는 위험을 갖고 있는가? 1-4. 기업은 개인정보를 활용하여 어떻게 이윤을 추구하는가?
		관심과 신뢰	1-5. 기관에서 제공하는 정보가 어떻게 이용, 저장, 공유되는지 사람들이 알고 있는가? 1-6. 자신에 관한 정보가 어떻게 이용, 저장, 공유되는지에 대하여 사람들은 어느 정도 관심을 갖는가? 다른 집단보다 특히 더 관심을 갖는 집단은 누구인가? 또는 다른 유형보다 더 관심을 얻는 데이터 유형은 무엇인가? 1-7. 어떠한 상업적 행동이 개인정보 이용에 관한 신뢰에 영향을 미치는가? 타겟팅 광고 또는 '추천'이 신뢰에 영향을 미치는가? 1-8. 개인정보보호법은 사람들이 개인정보 관리에 보다 관심을 갖도록 하였는가? 개인정보보호법은 사람들의 행동에 어떻게 영향을 미치는가? 1-9. 현행 법규정은 투명성과 신뢰 제고에 어느 정도 기여하고 있는가? 보다 신뢰를 향상하기 위하여 정부는 무엇을 시도해야 하는가?
	국제 비교	1-10. 개인에 관한 데이터 이용에 있어 국제적으로 비교할 수 있는 측정도구가 있는가? 어떤 도구가 가장 효과적인가?	

대영역	소영역	주제	질문
	참여	개인 정보 관리	2-1. 사람들은 자신의 개인정보를 어떻게 관리해야 하는지 알고 있는가? 자신의 개인정보에 대한 접근을 막을 수 있는 도구에 대해 알고 있는가?
		포용성	2-2. 데이터 이용과 참여에 있어, 지역별 또는 인구 집단별 차이가 있는가? 2-3. 데이터 이용과 참여에 있어, 분야별로 차이가 있는가? 2-4. 집단별로 데이터 이용과 참여에 있어 경험하는 장벽은 무엇인가? 소외 및 취약계층에게 이러한 장벽은 어떠한 영향을 미치는가? 2-5. 정부는 소외 및 취약계층이 데이터 이용과 참여에서 경험하는 장벽 제거를 위하여 무엇을 할 수 있는가? 디지털 사회의 어떤 요소들이 소외 및 취약계층에게 가장 유익하겠는가?
		데이터 기술	2-6. 오늘날 경제에서 기본적인 데이터 기술은 취업에 있어 얼마나 중요한가? 기본적 데이터 기술은 무엇인가? 2-7. 어떤 직업에서 데이터 기술의 중요성이 가장 큰가? 2-8. 관련 기술은 이용 및 지원 가능한가? 2-9. 관련 기술 또는 심화 기술은 교육 시스템을 통해 제공될 수 있는가?
경제	데이터 기반 경제 내 효과적 운영	경쟁	3-1. 중소기업 또는 비영리조직이 직면하는 어려움과 도전이 있는가? 이러한 문제는 조직의 유형에 따라 어떻게 다른가? 3-2. 중소기업은 대기업의 데이터 및 데이터 인프라에 어떻게, 어느정도 의존하는가? 3-3. 기업과 비영리조직의 참여에 대한 장벽을 제거하기 위해 정부가 개입해야 하는 것이 있는가? 어떤 종류의 개입이 필요하며, 이러한 개입은 분야별로 다른가?
		기술 발전	3-4. 중소기업과 비영리조직은 어떻게 그리고 어느정도 데이터 세계의 혁신에 기여하는가? 3-5. 중소기업과 비영리조직은 AI 또는 사물인터넷과 같은 새로운 기술의 유익을 충분히 누리고 있는가? 3-6. 기업은 미래기술의 발전이 데이터 이용방식을 어떻게 변화시킬 것으로 예상, 기대하는가?

대영역	소영역	주제	질문
효과적 데이터 이용을 통한 성장과 생산성 제고	효과적 데이터 이용을 통한 성장과 생산성 제고	생산성	4-1. 효율성 향상을 통해 기업 생산성을 견인할 수 있도록 데이터를 이용하는 방법은 무엇인가? 4-2. 데이터를 보다 효과적으로 이용함으로써 얻을 수 있는 생산성을 막는 장벽은 무엇인가? 4-3. 다른 조직이 배울 수 있는 우수 사례는 무엇인가? 4-4. 기업은 데이터 이용의 전문성을 어떻게 개발하는가?
		데이터 이용의 사회·환경적 이익	4-5. 데이터 이용은 기후변화나 생물 보안과 같은 핵심 이슈에 대하여 어떻게 환경적 이익을 제공할 수 있는가? 4-6. 데이터 이용은 고령화, 웰빙 등과 같은 핵심 이슈에 대하여 어떻게 사회, 보건적 이익을 제공할 수 있는가?
		데이터 접근성	4-7. 기업이나 비영리조직이 완전히 공개해야 하는 데이터는 무엇인가? 왜 그러한가? 4-8. 정부는 기업과 비영리조직이 보다 많은 데이터를 공개하도록 독려해야 하는가? 4-9. 정부는 기업과 비영리조직으로 하여금 보다 많은 데이터를 공유하도록 어떻게 독려할 수 있는가?
		공공 데이터를 통한 성장과 효율성	4-10. 공공 데이터를 통해 민간 부문의 성장을 촉진하고 공공의 (재무적, 비재무적) 이익을 확장할 수 있는 가장 좋은 방법은 무엇인가? 4-11. 공공 부문은 데이터 활용에 관한 규정을 결정하기 위하여, 이러한 이익을 어떻게 정량화 및 평가할 수 있는가?
		국제 비교	4-12. 데이터가 경제성장, 생산성 및 혁신을 어떻게 이끄는지에 관하여 국제적으로 비교할 수 있는 측정도구가 있는가? 어떤 도구가 가장 효과적인가?
정부	효과적 데이터 수집과 이용을 통해 정부 운영 및 서비스 향상	데이터 이용의 현재	5-1. 정부 및 확장된 공공 부문은 어떻게 데이터를 효과적으로 수집, 공유, 분석 및 저장하고 있는가? 이와 관련하여 좋은 사례와 그렇지 않은 사례는 어떤 것인가? 5-2. 정부 내에서 보다 효과적으로 데이터를 이용하는데 있어 주요한 장벽은 무엇인가? 정부와 민간 부문이 협업하는 경우에도 이러한 장벽이 있는가? 5-3. 정부 내 데이터 관리와 이용, 접근 등에 있어 기술 격차를 보이는 영역이 있는가? 정부에서 추가적으로 필요한 데이터 기술은 무엇인가? 5-4. 지방정부는 보유하고 있는 데이터를 어떻게 효과적으로 이용하는가? 어떤 도전과 어려움이 있는가? 5-5. 공공기관은 보유하고 있는 데이터를 어떻게 효과적으로 이용하는가? 어떤 도전과 어려움이 있는가?

대영역	소영역	주제	질문
		데이터 이용 향상	<p>5-6. 정부와 공공기관은 데이터 수집과 공유, 분석, 저장을 보다 개선하기 위하여 무엇을 할 수 있는가?</p> <p>5-7. 정부는 이러한 향상과 개선에 대하여 어떻게 인센티브를 창출할 수 있는가?</p> <p>5-8. 정부 내 데이터 이용과 관련하여 배울 수 있는 우수 사례는 무엇인가?</p> <p>5-9. 데이터 관리 및 거버넌스와 관련하여 공공부문이 배울 수 있는 민간 부문의 우수 사례는 무엇인가?</p> <p>5-10. 효과적인 데이터 이용을 통해 공공 서비스를 개선할 수 있는 주요 기회로는 무엇이 있는가?</p>
	협력적 데이터 이용을 통한 정부 조정 제고	데이터의 공유	<p>6-1. 공공기관은 그들이 보유하고 있는 데이터에 대한 접근을 언제 확대해야 하고, 언제는 그렇지 않아야 하는가?</p> <p>6-2. 사람들은 데이터 이용 개선을 통해 정부서비스가 향상되는 것을 언제 발견하는가? 사람들은 이러한 개선을 어떤 영역에서 가장 원하는가?</p> <p>6-3. 정부 내 데이터 공유 및 접근이 늘어남에 따른 이슈로는 무엇이 있는가? 프라이버시 향상 기술의 이용이 이러한 이슈를 해결할 수 있는가?</p> <p>6-4. 데이터 공유에 대한 장벽은 정부 및 공공기관에게 어떠한 문제를 야기하는가?</p> <p>6-5. 데이터 공유에 대한 장벽은 정부 외부 기관에 대하여 어떠한 문제를 야기하는가?</p> <p>6-6. 디지털경제법 제5부는 특정 목적을 위한 공공 정보의 이용에 대한 법적 근거를 제시한다. 이에 따라 조직들은 어떻게 데이터를 적절하고 효과적으로 이용할 수 있는가?</p>
		국제 비교	<p>6-7. 정부 데이터의 효과적 이용 및 그를 활용한 공공서비스 전달과 관련하여 국제적으로 비교할 수 있는 측정도구 및 사례가 있는가?</p>

* 출처: National Data Strategy 페이지(www.gov.uk/guidance/national-data-strategy)



* 출처: Roberts, S. (2019). How roundtables are helping inform the National Data Strategy. DCMS Blog (<http://dcmsblog.uk>)

그림 IV-2 국가 데이터 전략 수립을 위한 원탁회의

2) 빅데이터 관련 법제

빅데이터 관련 영국 법제로는 정부와 공공기관의 정보공개를 다루는 「Freedom of Information Act 2000(2000년 정보공개법)」과 개인정보 처리 및 보호에 관한 「Data Protection Act 2018(2018년 정보보호법)」, 그리고 연구자의 공공 데이터 접근에 대하여 법적 근거를 제공하는 「Digital Economy Act 2017(2017년 디지털경제법)」이 대표적이다.

① Freedom of Information Act 2000(2000년 정보공개법)

먼저 Freedom of Information Act 2000(2000년 정보공개법)은 공공기관 정보에 대한 국민의 접근권, 궁극적으로 알 권리를 보장하기 위한 법으로 2000년

11월 30일 제정되었다(영국의 법률데이터베이스). 2000년 정보공개법은 공공기관에게 주요 활동에 관한 정보들을 공개할 책임을 부여하는 동시에 국민으로 하여금 공공기관에게 정보를 요구할 수 있는 권리를 보장하는 것을 골자로 하는데, 이때 공공기관은 잉글랜드와 웨일즈, 북아일랜드 지역의 모든 공공기관 및 스코틀랜드 지역에 위치한 영국 연방정부 기관 즉 정부 부처, 지방정부, 국가보건의료서비스(NHS), 공립학교, 경찰청 등을 모두 포함하는데, 스코틀랜드 지역 공공기관은 스코틀랜드 지역에서 제정한 「Freedom of Information (Scotland) Act 2002(2002년 스코틀랜드 정보공개법)」를 따른다. 이 법에서 말하는 정보는 인쇄물, 컴퓨터 파일, 서신과 이메일, 사진, 녹음 및 영상 기록 등을 모두 포함하되 건강기록이나 신용정보와 같은 개인정보는 포함되지 않는다.

2000년 정보공개법은 총 8부 88개 조항으로 구성되어 있는데, 제1부(제1조-제20조)는 공공기관 정보에 대한 접근권을 주제로 정보에 대한 권리, 정보공개를 거절할 수 있는 경우, 정보위원회, 정보공개(발행) 계획 등에 대하여 규정하고 제2부(제21조-제44조)는 정보공개 의무에서 면제되는 경우 예를 들어 국가안보 및 방위 관련 정보, 국제관계 관련 정보 등 예외 경우에 대하여 설명한다. 제3부(제45조-제49조)는 국무조정실장, 내각 장관 및 정보감독관(Information Commissioner)의 역할, 제4부(제50조-제56조)는 법률집행, 제5부(제57조-제61조)는 항소, 제6부(제62조-제66조)는 공공기록사무소 내 역사적 기록물에 관한 사항, 제7부(제68조-제73조)는 개인정보 보호 관련 사항, 제8부(제74조-제88조)는 그 외 기타사항을 규정한다.

② Data Protection Act 2018(2018년 정보보호법)

Data Protection Act 2018(2018년 정보보호법)은 개인정보처리 및 보호에 관한 규정으로 2018년 5월 23일 제정되었다(영국의 법률데이터베이스). EU에서 개인정보 보호를 위한 법령으로 'General Data Protection Regulation(GDPR)'

을 2016년 제정(2018년 5월 시행)함에 따라 마련되었는데 개인정보의 공정하고 합법적이며 투명한 이용, 구체적이고 명확한 이용 목적, 적절하고 이용 목적에 제한되는 이용 방법 등의 개인정보 보호 원칙을 규정하고 민감한 정보(예: 인종, 윤리적 배경, 정치적 의견, 종교적 신념, 생체정보, 건강, 성적 지향 등)에 대한 보다 강력한 법적 보호를 명하고 있다. 이를 통해 국민은 누구나 정부, 공공 또는 민간 기관에서 자신의 정보를 수집, 저장, 이용, 삭제하는 것에 대한 알 권리, 부정확한 정보에 대하여 수정을 요청할 권리, 정보처리에 대한 중지 및 금지를 요청할 권리 등을 보장받는다.

구체적으로 2018년 정보보호법은 총 7부 215개 조항으로 이루어져 있는데, 제1부(제1조-제3조)는 서문으로 본 법의 개관 및 주요 용어 개념 정의 등을 다루고, 제2부(제4조-제28조)는 일반적 개인정보보호규정, 구체적으로 개인정보보호 규정이 적용되는 범위, 주요 용어의 정의, 개인정보의 범주, 정보주체의 권리 및 정보주체의 권리가 제한되는 경우 등, 제3부(제29조-제81조)는 개인정보보호 법률집행, 제4부(제82조-제113조)는 정보기관, 제5부(제114조-제141조)는 정보 보호감독관(Information Commissioner), 제6부(제142조-181조)는 법률집행, 제7부(제182조-제215조)는 추가 조항을 규정한다.

③ Digital Economy Act 2017(2017년 디지털경제법)

Digital Economy Act 2017(2017년 디지털경제법)은 디지털 커뮤니케이션 인프라와 서비스에 관한 법률로 2016년 7월 법안이 발의되어 2017년 4월 17일 제정되었다(영국의 법률데이터베이스). 총 7부 120개 조항으로 이루어져 있는데 제1부(제1조-제3조)는 디지털 서비스에 대한 접근, 제2부(제4조-제13조)는 디지털 인프라, 제3부(제14조-제30조)는 온라인 포르노그래피, 제4부(제31조-제34조)는 지적재산권, 제5부(제35조-제81조)는 전자정부, 제6부(제82조-제116조)는 기타 조항, 제7부(제117조-제120조)는 일반 조항을 규정한다.

이때 제5부 전자정부 중 제5장이 연구 목적을 위한 공유(Chapter 5: Sharing for research purposes)를 다루어 연구자들의 공공데이터 접근에 대한 법적 근거를 제시한다. 특히 제64조는 연구 목적을 위한 정보공개를 규정하는데, 공공기관이 공적 기능과 관련하여 보유하는 정보를 익명화하여 특정 개인을 식별, 지정할 수 없을 때 그 정보를 연구 목적으로 다른 사람에게 제공할 수 있음을 밝히고 있다. 이를 통해 연구자들이 공공데이터(빅데이터)에 안전하고 윤리적 방식으로 접근하여 그에 기반한 정책 연구를 수행할 수 있는 법적 근거가 제공되었다 할 수 있다.

3) 빅데이터 정책 추진 체계

빅데이터 관련 정책의 주요 추진 체계로 먼저 정부의 빅데이터 정책의 주관 부처라 할 수 있는 ‘Department for Digital, Culture, Media & Sport(디지털·문화·미디어·스포츠부, 이하 “DCMS”)', 그리고 빅데이터 정책의 직접적 주관 부처는 아니나 내각사무처(Cabinet Office)의 산하 조직으로 정부의 디지털화를 담당하는 ‘Government Digital Services(정부디지털서비스청, 이하 “GDS”)', 정부의 정보공개 및 개인정보 보호에 관한 독립적 감독기구라 할 수 있는 ‘Information Commissioner’s Office(개인정보감독기구, 이하 “ICO”)', 열린 정책 수립을 위한 기구로 여러 정책 실험을 추진하는 가운데 ‘국가 데이터 전략’ 수립 과정에도 함께 하고 있는 ‘Policy Lab(정책실험실)’ 그리고 오픈 데이터 관련 연구소로는 세계 최초로 설립된 ‘Open Data Institute(오픈 데이터 연구소, 이하 “ODI”)' 등을 고려할 수 있다.

① Department for Digital, Culture, Media & Sport(DCMS, 디지털·문화·미디어·스포츠부)

DCMS는 영국의 빅데이터 관련 정책을 추진하는 핵심 부처로 문화유산, 예술 등에서부터 인공지능에 이르기까지 디지털, 문화, 미디어, 스포츠와 관련한 다양한 분야를 주관한다(영국 정부포털). 현재 올리버 다우든 장관(Rt Hon. Oliver Dowden)이 이끌고 있으며 약 1,000여명의 직원이 소속되어 있다. 빅데이터와 관련하여 「UK Digital Strategy(영국 디지털 전략, 2017년 발표)」, 「National Data Strategy(국가 데이터 전략, 2020년 발표 예정)」 등 정부의 주요 정책의 계획과 추진, 총괄 관리 등의 역할을 맡고 있다.

② Government Digital Service(GDS, 정부디지털서비스청)

GDS는 내각사무처의 산하조직으로 정부서비스의 디지털화 및 혁신을 통해 정부와 국민 간 상호작용 및 정부 활동의 효과성과 효율성을 증진하는 것을 목적으로 한다(영국 정부포털). GDS는 2012년 ‘Government Digital Strategy(정부 디지털 전략)’ 추진을 담당하였고 2017년부터는 ‘Government Transformation Strategy(정부 혁신 전략)’를 추진하는 중이다. 영국은 전 세계적으로 우수한 전자정부 시스템을 갖춘 국가 중 하나로, UN(2020)이 최근에 발표한 ‘United Nations E-Government Survey 2020’에서 조사 참여국 192개국 중에서 7위를 차지(EGDI=0.9358)한 바 있다. 이처럼 GDS는 빅데이터 관련 정책을 직접 주관하고 있지는 않으나, 정부의 정보공개 및 디지털 혁신 추구를 담당하고 있기에 주요 추진 체계의 하나로 볼 수 있다.

③ Information Commissioner’s Office(ICO, 개인정보감독기구)

ICO는 DCMS 산하기구 중 하나이지만 동시에 정보 보호 및 정보에 관한 권리 증진을 위한 독립기구로, ‘디지털 시대에서 영국 국민의 정보기본권을 수호하는

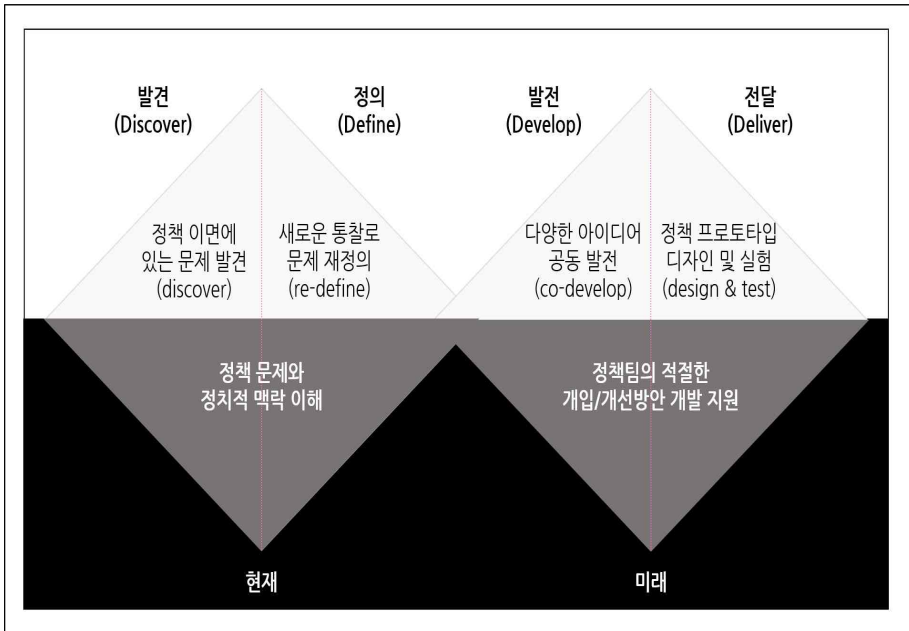
것(To uphold information rights for the UK public in the digital age)'을 사명으로 한다(Information Commissioner's Office, 2017). 이러한 사명 아래 ICO는 (i) 정보보호법의 효과적 적용과 관리감독, (ii) 프라이버시 보호를 위한 보다 혁신적 방안 모색, (iii) 정보와 관련하여 투명성과 책무성 제고 및 정보 거버넌스 증진, 그리고 궁극적으로 (iv) 디지털 세계에서 국민 보호를 추구하고 있다. ICO는 빅데이터 관련 정책의 기획과 실행을 담당하는 기구는 아니지만, 빅데이터 관련 핵심 법제라 할 수 있는 2000년 정보공개법과 2018년 정보보호법에 따른 관리감독 역할을 담당하고 있는 곳이기에 중요한 체계 중 하나로 볼 수 있을 것이다.

④ Policy Lab(정책실험실)

Policy Lab은 2014년 '열린 정책 수립(Open Policy Making)'을 목표로 공공 서비스를 개혁하려는 계획에 따라 영국 내각사무처에 설치된 정부 기구다(영국 정부포털 내 열린 정책 블로그). 여러 정부 부처와의 협력 아래 데이터, 디자인, 디지털 도구 등을 활용하여 많은 정책 실험을 해오고 있으며, 지난 2019년부터는 '국가 데이터 전략(National Data Strategy)' 수립 과정에 참여하며 보다 나은 데이터 정책 수립을 위해 노력 중이다. Policy Lab은 국가 데이터 전략 수립을 위해 관련 근거자료를 수집·분석하고 그 결과를 토대로 전국에서 20개 이상의 원탁회의를 이끌고 있는데, 원탁회의에는 학계와 시민사회, 기업, 공공기관 등에서 250여 단체들이 참여하고 있다.

Policy Lab은 정책 설계 및 실험은 Design Council(디자인 위원회)의 디자인 씽킹 프레임워크 곧 '발견(Discover) - 정의(Define) - 발전(Develop) - 전달(Deliver)'의 과정을 따른다(영국 정부포털 내 열린 정책(Open Policy) 블로그). 구체적으로 정책 이면에 있는 문제를 발견(discover)하고 재정의(re-define)하며 문제와 그를 둘러싼 정치적 맥락을 이해하고, 이를 해결하기 위한 다양한 아이디어를 발전시키고(co-develop) 정책으로 디자인하고 실험함으로써(deliver) 관련

정책팀을 지원하는데, 아래 그림으로 Policy Lab의 정책 디자인 및 실험 과정을 요약하였다. 2019년부터 참여 중인 ‘국가 데이터 전략’ 또한 이러한 과정을 거쳐 정책 고안과 실험이 진행 중이라 할 수 있다.



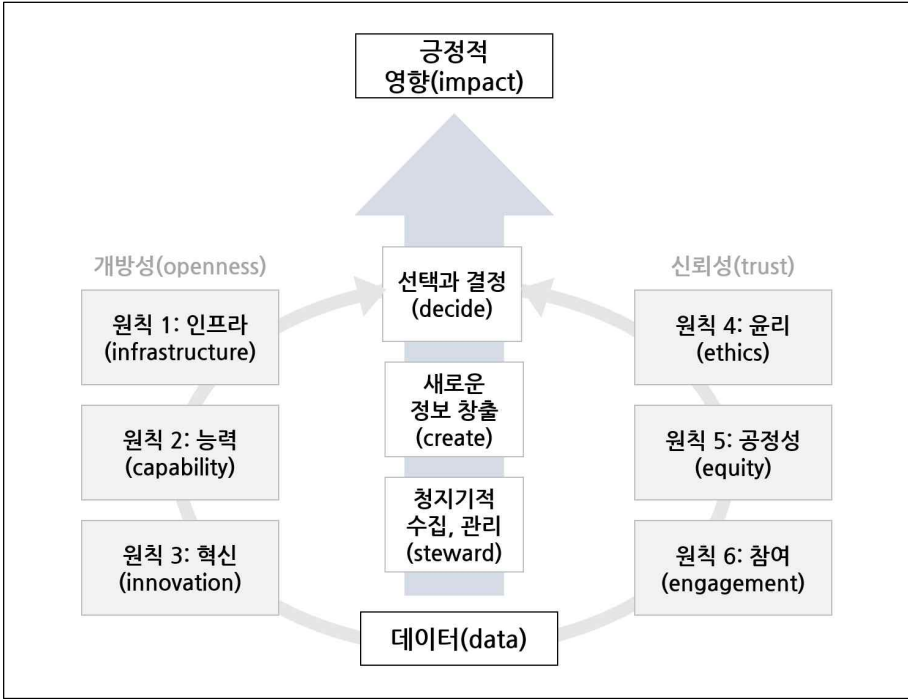
* 출처: 열린 정책 블로그(<https://openpolicy.blog.gov.uk>)에서 제시하는 Policy Lab 정책 디자인 프레임워크를 번역하여 재구성.

그림 IV-3 Policy Lab의 정책 디자인 프레임워크

⑤ Open Data Institute(ODI, 오픈 데이터 연구소)

‘Open Data Institute(ODI, 오픈 데이터 연구소)’는 2012년 월드 와이드 웹(world wide web, www)의 하이퍼텍스트 시스템을 개발하여 ‘웹의 아버지’로 불리는 컴퓨터 과학자, Tim Berners-Lee 경과 인공지능 전문가 Nigel Shadbolt 경이 공동으로 설립한 비영리 연구기관으로 데이터를 활용한 혁신을 위해 전세계 사람들을 연결하고 준비시키고 고취시키는 것(to connect, equip and inspire people around the world to innovate with data)을 미션으로 한다(Open Data Institute 홈페이지).

특히 ODI는 데이터를 통한 혁신이 사회에 미치는 긍정적 영향을 아래 그림과 같이 ‘변화이론(Theory of Change)’으로 설명한다. 즉, 데이터를 청지기적으로 수집, 보관, 관리하고(stewarding data), 데이터에 기반하여 새로운 정보를 창출하고(creating information from the data), 그러한 정보들을 토대로 선택과 결정을 내림으로써(deciding what to do), 모두를 위한 긍정적 영향을 만들어낼 수 있는데 이때 데이터의 청지기적 관리, 정보 창출과 그에 따른 결정 과정은 인프라, 능력, 혁신, 윤리, 공정성, 참여의 6대 원칙 아래 이루어져야 한다는 것이다(Open Data Institute 홈페이지). 아래 그림으로 변화이론의 주요 내용을 제시하였다.



* 출처: 오픈 데이터 연구소 홈페이지(<http://theodi.org>)에서 제시하는 변화 이론 내용을 번역하여 재구성.

그림 IV-4 ODI의 변화 이론(Theory of Change)

4) 빅데이터 기반 정책 연구 체계: Administrative Data Research UK (ADR UK)

영국의 빅데이터 기반 정책 연구 체계인 Administrative Data Research UK(ADR UK)는 연구자들이 공공 데이터를 활용하여 과학적 연구를 수행하여 증거를 생산, 이를 통해 정부의 정책 설계 또는 결정에 기여하는 것을 목적으로 한다. ADR UK의 뿌리는 2012년 12월 출범한 ‘Administrative Data Research Network(ADRN)’으로, 2012년 당시 영국 ‘경제사회연구위원회(Economic and Social Research Council)’가 공공 데이터를 연구 목적으로 활용하는 사업을 추진하고자 재무부로부터 예산을 지원받아 대학, 정부, 산업계의 전문가들로

태스크포스(Taskforce)를 꾸린 것에서 출발한다(영국 통계청 홈페이지). ADRN은 잉글랜드, 웨일즈, 스코틀랜드, 북아일랜드 4개 권역에 ‘공공 데이터 연구센터(Administrative Data Research Centre, ADRC)’ 및 데이터 접근을 위한 단일 창구로서 ‘공공 데이터 서비스(Administrative Data Service)’를 설립하였고 영국 통계청(UK Statistics Authority, UKSA)이 ADRN에 대한 의회 보고의 책임을 졌다.

이후 2018년 ADRN은 ADR UK로 발전하였다. ADR UK는 ‘공공 데이터에 대한 연구자의 접근 방식을 혁신한다’는 미션 및 아래의 4가지 핵심 가치 아래, 공공 데이터의 연구 활용을 지원, 관리함으로써 보다 나은 사회로의 변화에 기여하고자 한다(ADR UK, 2019).

책임 있고 윤리적인 데이터 활용(Responsible and ethical data use): ADR UK는 공공 데이터가 사회에 긍정적으로 기여할 수 있는 연구에 대하여 책임 있고 윤리적으로 활용될 것을 보장한다. 이를 위하여 데이터는 익명화되어야 하고 엄격하게 통제 및 모니터링되어야 하며, 윤리와 타당성, 프라이버시 보호, 공익, 과학적 유용의 관점에서 끊임없이 평가되어야 한다.

독립성(Independence): ADR UK는 중립적이고 독립적 기구이며, 공공 데이터를 활용하는 연구는 특정 아젠다를 추구하기보다는 사회를 보다 더 잘 이해하고 보다 나은 사회로 개선되도록 하고자 하는 열망에 따른 학문 연구여야 한다.

연관성(Relevance): 공공 데이터를 활용하는 연구는 사회의 필요와 연관되어야 한다. ADR UK는 영국 사회의 중대한 이슈와 연구주제의 연관성, 공익에 대한 기여 가능성 등을 기준으로 연구 제안서를 철저히 검토한다.

미래를 향한 큰 열망(Ambition): ADR UK는 공공 데이터를 활용한 연구 수행을 통한 사회적 변화 그리고 그러한 연구를 위한 안전한 데이터 공유가 일상화되고 존중되는 미래사회를 향한 비전을 갖고 있다.

ADR UK의 파트너십은 ADR 스코틀랜드와 ADR 웨일즈, ADR 북아일랜드 그리고 국가통계사무소(Office for National Statistics, ONS)로 구성되고 ADR

UK Strategic Hub(전략 허브)에 의해 조정·관리되며, ADR UK 내부는 프로그램 전달 위원회(Programme Delivery Board, PDB), 리더십 위원회(Leadership Committee, LC), 운영관리 그룹(Operational Management Group, OMG) 그리고 연구위원회(Research Commissioning Board, RCB)로 구성된다(ADR UK, 2019). 프로그램 전달 위원회(PDB)는 ADR UK의 전략적 목표 이행의 제반 과정을 관리·감시하고 리더십 위원회(LC)는 ADR UK의 파트너 기관들이 ADR UK의 전략적 방향을 동의하고 발전시키도록 하며 운영관리 그룹(Operational Management Group, OMG)은 ADR UK의 운영을 점검, 보고하기 위한 포럼으로 파트너 기관들의 이행 성과를 자체 평가, 관리하기 위한 증거를 수집하는 역할을 담당한다. 연구위원회(RCB)는 공공데이터를 활용하고자 하는 연구 제안서가 공익에 기여하는지 검토, 평가하여 새로운 연구를 발굴, 승인하는 역할을 담당하는데 RCB 구성원은 모두 ADR UK 외부 인사로 구성된다.

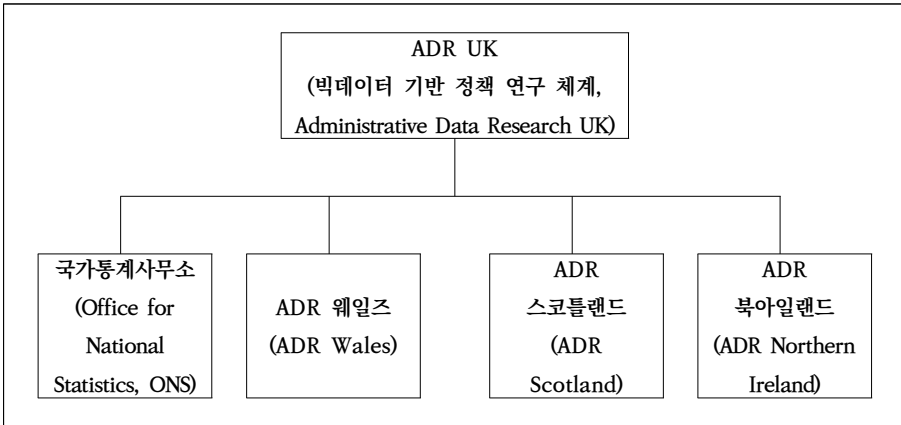


그림 IV-5 ADR UK의 파트너십

* 출처: ADR UK (2019). Administrative data is an invaluable resource for public good. Let's use it (Annual Report 2018-2019): 12.

ADR UK는 영국의 빅데이터 기반 정책 연구 체계로서 공공데이터를 활용한 연구 프로젝트 수행을 통해 정부 부처 및 공공기관의 정책 설계를 위한 기초자료를 제공하고자 한다. ADR UK의 빅데이터 기반의 연구 프로젝트는 주택과 커뮤니티, 건강과 웰빙, 아동·청소년, 직업 세계, 고령화, 불평등과 사회통합, 기후위기와 지속가능성, 범죄와 사법의 8개 주제 아래 총 58개 프로젝트가 수행 중이다(ADR UK 홈페이지). 아래 표로 주제별 대표적 연구 프로젝트를 제시하였다.

표 IV-7 ADR UK의 연구주제 및 프로젝트

연구주제	프로젝트
주택과 커뮤니티	<ul style="list-style-type: none"> • 주택과 웰빙의 관계 탐색(Exploring the relationship between housing and wellbeing) • 잉글랜드 노숙인 데이터: 노숙인 정책을 평가하기 위한 지방정부 데이터 연결(Homelessness Data England: Linking local authority data to evaluate homelessness policy)
건강과 웰빙	<ul style="list-style-type: none"> • 청각손실과 치매의 관계 규정(Identifying the links between hearing loss and dementia) • ADR 웨일즈의 Covid-19 대응(ADR Wales Covid-19 Response)
아동·청소년	<ul style="list-style-type: none"> • 사회적 배제와 교육성취의 관계(The relationship between social exclusion and educational attainment) • GUIE 데이터 세트 구성과 테스트: 유랑민족(RGT) 자녀 사례 연구(Shaping and testing the GUIE dataset: Roma, Gypsy and Traveller children case study)
직업 세계	<ul style="list-style-type: none"> • 웨일즈 지역에서 ESF의 지원을 받는 노동시장 프로그램의 효과성 검증(Assessing the effectiveness of ESF-funded labour market programmes in Wales) • 고등교육 참여에 대한 진로지도의 역할(The role of careers guidance in supporting participation in higher education)
고령화	<ul style="list-style-type: none"> • 할인운임 이용자 이해(Understanding users of concessionary fares) • 노년기 돌봄(Care in the last years of life)
불평등과 사회통합	<ul style="list-style-type: none"> • 노숙문제 해결을 위한 방법론적 접근(Methodological approach to tackling homelessness) • 북아일랜드의 약물 관련 사망: 사회인구학적 분석(Drug-related deaths in Northern Ireland: Socio-demographic analyses)

연구주제	프로젝트
기후위기와 지속가능성	<ul style="list-style-type: none"> 건강과 생명에 대한 대기오염의 영향(The effect of air pollution on health and mortality) 어머니의 대기오염 노출과 신생아의 건강(Infant health and mothers' exposure to air pollution)
범죄와 사법	<ul style="list-style-type: none"> 불법 약물의 위탁(Illegal drug consignments) 폭력 관련 사건 및 그 아래에 있는 취약성(Violence-related incidents and underlying vulnerability)

* 출처: ADR UK 홈페이지(<https://www.adruk.org>).

5) 빅데이터 기반 아동·청소년 연구 사례: ADR UK의 아동·청소년 연구 프로젝트

영국의 빅데이터 기반 정책 연구 체계인 ADR UK(Administrative Data Research UK)는 아동·청소년에 관한 빅데이터 기반 연구 프로젝트로 총 26개⁵⁹⁾를 승인, 지원하고 있다. 연구 프로젝트는 주로 건강 및 정신건강, 학업 및 학교생활, 요보호 또는 취약아동, 아동·청소년 관련 정부 개입과 지원 등을 다루는데, 다음의 표로 각 연구 프로젝트의 주제와 지원 기관을 제시하였다.

표 IV-8 ADR UK가 승인, 지원하고 있는 아동·청소년 관련 연구 프로젝트

연번	연구 프로젝트	지원 기관
1	GUIE(Growing Up in England) 데이터 세트 구성과 테스트: 유랑민족(RGT) 자녀 사례 연구(Shaping and testing the GUIE dataset: Roma, Gypsy and Traveller children case study)	ADR UK 전략 허브
2	사회적 배제와 교육성취의 관계(The relationship between social exclusion and educational attainment)	ADR 북아일랜드
3	어머니의 대기오염 노출과 신생아의 건강(Infant health and mothers' exposure to air pollution)	ADR 북아일랜드

59) 하나의 연구 프로젝트가 2개 이상의 주제를 포함할 수 있으며, 아동·청소년에 관한 26개 연구 프로젝트는 각 프로젝트 주제 중 '아동·청소년'이 포함되는 모든 연구를 의미한다.

연번	연구 프로젝트	지원 기관
4	'Pupil Deprivation Grant(PDG, 취약학생 지원기금의 하나로 무료급식 및 돌봄서비스 제공)'와 학업성취 격차의 관계(School-level attainment gaps and the Pupil Deprivation Grant)	ADR 웨일즈
5	교육에 대한 노숙의 영향(The impact of homelessness on education)	ADR 웨일즈
6	관련 데이터를 활용한 학교 준비도 예측(Predicting school readiness using linked data)	ADR 웨일즈
7	학업성취와 빈곤 극복의 관계(The relationship between educational attainment and surviving poverty)	ADR 웨일즈
8	임신 기간 중 음주가 아동에 대하여 미치는 장기적 영향(Can mild alcohol use during pregnancy affect children in the long-term?)	ADR 웨일즈
9	사춘기 때 이른 음주 경험: 기록 및 데이터 기반의 종단 분석 연구(Early alcohol use in adolescence: A record linked, data-driven, predictive analytical longitudinal study)	ADR 웨일즈
10	아동·청소년의 정신건강 장애와 교육유형 간의 관계(Association between mental health disorders and patterns of education in children & young people)	ADR 웨일즈
11	정신건강 장애와 교육유형 간의 관계 탐색(Exploring associations between mental health disorders and patterns of education)	ADR 웨일즈
12	웨일즈 학교 기반 상담 전략의 영향 탐색(Exploring the impact of the Welsh school-based counselling strategy)	ADR 웨일즈
13	북아일랜드의 정책 변화 이후 치과 및 교정 서비스 접근(Access to dental and orthodontic services in Northern Ireland after policy change)	ADR 북아일랜드
14	대학생의 정신건강과 자해 행동: 위험요인과 궤적(Mental health and self-harming behaviours in university students: Risk factors and trajectories)	ADR 웨일즈
15	의무교육 기간 이후의 교육과 훈련 참여 이해(Understanding participation in post-compulsory education and training)	ADR 웨일즈
16	고등교육 참여에 대한 진로지도의 역할(The role of careers guidance in supporting participation in higher education)	ADR 웨일즈
17	잉글랜드 아동의 건강과 교육에 관한 데이터 연령별 집합(The All Years Dataset: Linking children's health and education data for England)	ADR UK 전략 허브

연번	연구 프로젝트	지원 기관
18	스코틀랜드 지역 아동의 삶과 성과 이해(Understanding Children's Lives and Outcomes in Scotland)	ADR 스코틀랜드
19	GUIE(Growing Up in England) 데이터 집합: 2011 인구센서스 및 아동 교육 데이터(Growing Up in England: Linking 2011 Census and education data for children)	ADR UK 전략 허브
20	폭력 및 범죄예방을 위한 정부의 개입 활동의 평가를 위한 법무부와 교육부 데이터 활용 가능성 연구(Feasibility study to explore the use of the MoJ-DfE linked dataset to aid evaluations)	내무부
21	북아일랜드 지역의 돌봄 아동에 대한 지원 성과(Outcomes for looked-after children in Northern Ireland)	ADR 북아일랜드
22	긍정적 청소년 발달 조사(Investigating positive youth development)	ADR 스코틀랜드
23	중등교육 참여와 장기적 건강의 관계(Selective schooling and long-term health)	ADR 스코틀랜드
24	가정 외 보호 아동의 위탁 보호 안정성(Placement stability of children in out-of-home care)	ADR 스코틀랜드
25	청소년 정신건강에 영향을 미치는 요인 연구(Examining factors relating to adolescent mental health)	ADR 북아일랜드
26	스코틀랜드 지역 돌봄 아동의 건강 욕구 충족(Meeting the health needs of looked-after children in Scotland)	ADR 북아일랜드

* 출처: ADR UK 홈페이지(<https://www.adruk.org>).

ADR UK이 승인, 지원 중인 아동·청소년 연구 프로젝트 중 건강 관련 연구 사례로 ‘어머니의 대기오염 노출과 신생아의 건강(Infant health and mothers’ exposure to air pollution)’, 학업 관련 연구 사례로 ‘사회적 배제와 학업성취의 관계(The relationship between social exclusion and educational attainment)’, 요보호 및 취약아동 관련 연구 사례로 ‘가정 외 보호 아동의 위탁 보호 안정성(Placement stability of children in out-of-home care)’, 정부의 개입 및 지원 프로그램 효과성에 관한 연구 사례로 ‘웨일즈 학교 기반 상담 전략의 영향 탐색(Exploring the impact of the Welsh school-based counselling strategy)’을 선정하였다. 각 연구 프로젝트의 세부 내용은 다음과 같다.

① 아동·청소년 건강 관련 연구 사례: 어머니의 대기오염 노출과 신생아의 건강
(Infant health and mothers' exposure to air pollution)

‘어머니의 대기오염 노출과 신생아의 건강’ 연구 프로젝트는 북아일랜드 지역 여성의 임신기간 중 대기오염 노출이 출산 결과에 미치는 영향을 조사하는 것을 목적으로 한다. 본 연구의 총괄책임자는 벨파스트 퀸즈대학교(Queen's University Belfast)의 Duncan McVicar 교수이며, ADR UK의 주요 파트너 중 하나인 ADR 북아일랜드가 연구비를 지원한다.

연구에서 사용하는 데이터는 북아일랜드 모성서비스(Northern Ireland Maternity Service, NIMATs)에서 제공하는 데이터와 대중에게 공개되는 환경오염 데이터다. NIMATs는 어머니와 신생아에 관한 인구학적 정보와 건강 정보, 예를 들어 신생아의 출생 시 몸무게, 어머니의 과거 병력과 산과력 등 광범위한 정보를 담고 있다. 대기오염 노출 정보는 어머니가 임신 시기에 거주하였던 지역의 연평균 대기오염 수치로 산출하는데, 1x1 km² 격자 단위를 사용하여 계산한다. 연구는 임신기간 동안 상대적으로 대기오염이 심각한 지역에 거주한 어머니에게서 태어난 신생아가 일찍 태어나거나 또는 출생 시 체중이 적게 나가는지 등을 조사한다. 대기질 향상과 출산률 제고 모두 북아일랜드 정부의 주요한 정책 목표에 해당하는데, 이 프로젝트는 북아일랜드 정책 입안자들로 하여금 지역 내 신생아 건강 향상을 위해 대기오염 규제를 어느 정도로, 어떻게 해야 할지에 대한 의미 있는 정보를 제공해 줄 수 있다.

② 아동·청소년 학업 관련 연구 사례: 사회적 배제와 학업성취의 관계(The relationship between social exclusion and educational attainment)

‘사회적 배제와 학업성취의 관계’ 연구 프로젝트는 학교로부터 일시적 또는 영구적 배제를 경험한 학생이 보다 낮은 학업성취를 보이는지 검증하는 것을 목적으로 한다. 구체적으로 정학, 퇴학 등의 배제 유형, 배제 사유 그리고 성별이나 연령

과 같은 학생 개인적 특성이 학업성취에 미치는 영향을 조사하고자 계획 중이다. 앞의 연구와 마찬가지로 벨파스트 퀸즈대학교(Queen's University Belfast)의 Duncan McVicar 교수가 연구의 총괄책임을 맡고, ADR 북아일랜드가 연구비를 지원한다.

연구에서 사용하는 데이터는 전국 학생 데이터베이스(National Pupil Database, NPD)다. NPD는 잉글랜드 지역의 초등학교에서 대학에 이르는 교육 기관에 재학 중인 학생 전체에 대하여 상당히 광범위한 정보를 포함한다. 예를 들면 학생의 인구학적 정보, 학교 특성에 관한 다양한 정보, 학생의 결석, 정학 및 퇴학 이력, 학생의 학업성취 기록(Key Stage 1 즉 7세부터 Key Stage 5 즉 18세까지의 학업성취 기록) 모두가 수록되어 있다. 이러한 사회적 배제 즉 정학 또는 퇴학 경험과 학업성취 관계를 검증함으로써 학생에 대한 학교, 나아가 가족과 지역사회의 적절한 개입과 지원을 논할 수 있는 단초를 제공할 수 있다는 점에 본 연구 프로젝트의 의의가 크다 하겠다.

③ 요보호 및 취약아동 관련 연구 사례: 가정 외 보호 아동의 위탁 보호 안정성
(Placement stability of children in out-of-home care)

‘가정 외 보호 아동의 위탁 보호 안정성’ 연구 프로젝트는 스코틀랜드 지역의 가정 외 보호 아동의 위탁 보호의 중단적 결과를 세밀하게 분석하는 것을 목적으로 한다. 연구 책임자는 에딘버러 대학의 Ben Matthews 박사와 Gillian Raab 박사이며, ADR UK의 주요 파트너 중 하나인 ADR 스코틀랜드가 연구비를 지원한다.

연구에서 사용하는 데이터는 스코틀랜드 학생 센서스(Scottish Pupil Census)와 스코틀랜드 지역 위탁보호 아동 데이터(Looked-After Children data)로, 2008년부터 2015년의 7년간에 대한 자료이다. 이 연구는 2008년 데이터를 토대로 위탁 보호 아동의 코호트를 구축하고 이들의 발달 궤적을 추적함으로써 7년간 이들이 경험한 위탁보호 유형 및 안정성, 즉 어떠한 곳에서 어떠한 보호를 경험하

였는지 그리고 한 가정에서 위탁보호를 받았는지 또는 여러 가정을 이동하였는지 등의 안정성을 조사한다. 이러한 위탁 보호 경험과 안정성이 아동의 학교생활 곧 출결석 및 퇴학 등에 영향을 미쳤는지 살펴보고자 스코틀랜드 정부 출결석 데이터(Scottish Government's Attendance and Absence in Scottish Schools)와 퇴학 데이터(Scottish Government's Schools Exclusions data) 등을 함께 이용할 예정이다.

④ 개입 프로그램의 효과성 관련 연구 사례: 웨일즈 학교 기반 상담 전략의 영향 탐색(Exploring the impact of the Welsh school-based counselling strategy)

‘웨일즈 학교 기반 상담 전략의 영향 탐색’ 연구 프로젝트는 웨일즈 정부가 중등학교를 통해 운영하는 상담 서비스의 효과, 주로 우울과 불안, 항정신성 약물 등에 대한 영향을 검증하는 것을 목적으로 한다. 웨일즈 지역의 스완지 대학(Swansea University)의 Ann John 교수와 Jo McGregor 교수가 연구의 총괄책임 맡고, ADR 웨일즈가 연구비를 지원한다.

웨일즈 정부는 정서적, 행동적 또는 사회적 문제가 있는 아동·청소년에게 효과적인 초기 개입으로서 상담의 가치와 영향력에 주목하며 지역의 모든 중등학교 내 상담 서비스를 도입, 운영하고 있으며, 이러한 학교 기반 상담 전략의 효과성을 검증하고자 2010년 영국상담정신치료협회(British Association for Counselling and Psychotherapy), 스트래스클라이드 대학교(University of Strathclyde)와 공동으로 평가를 실시한 바 있다. 그 결과 학교 기반 상담 전략 도입 및 그에 따른 상담 서비스가 내담자를 포함한 모든 이해관계자들에게 효과적, 성공적으로 평가되고 있는 것이 확인되었다.

웨일즈 정부는 그에서 한 걸음 더 나아가, 이번 빅데이터 기반 연구 프로젝트를 통하여 학교 기반 상담 전략이 실제 학생들의 정신건강에 기여하는지, 예를 들어

우울이나 불안 등의 문제 완화에 영향을 미치는지 검증함으로써 학교 기반 상담 전략에 대한 재정 지원 증대 및 초등학교 도입 추진을 위한 객관적 근거를 확보하고자 한다.

6) 시사점

① 정부 주도의 데이터 정책 추진

영국은 일찍부터 데이터의 가치에 주목하며 정부 주도의 데이터 정책을 추진해 왔다. 2000년 정보공개법 제정을 시작으로 2009년 「스마트 정부 구현을 위한 실행계획서」, 2012년 「오픈 데이터 백서」, 2014년 「오픈 데이터 전략 2014-2016」 등을 발표하며 공공 데이터 개방 정책을 펼치며 정부 운영의 효율성과 투명성, 궁극적으로 책무성을 제고하고자 하였다. 이후에는 데이터 분석과 활용에 집중하며 2013년 범국가적 「데이터 역량 강화 전략」, 2017년 「디지털 전략」 등을 추진하였고 현재는 「국가 데이터 전략」을 고안하는 중이다. 이처럼 영국 정부는 (빅)데이터를 미래 영국을 이끌어갈 주요한 성장동력으로 여기며, 관련 법제를 마련하고 정책을 추진하는데 매우 적극적인 모습이다.

이러한 정부의 높은 관심은 빅데이터 관련 법제적, 행정적 준비를 촉진할 뿐 아니라 학계와 기업, 시민단체 등 민간 부문의 참여와 협력을 보다 강력하게 견인할 수 있기에 중요하다. 세계은행은 빅데이터 시대에서 방대한 공공 데이터를 보유하고 있는 정부는 데이터 공급자로서 중요하며, 뿐만 아니라 다른 공공기관, 민간기업, 시민단체 등의 데이터 개방을 위한 선도자이자 촉진자의 역할도 담당할 수 있다며 그 중요성을 강조한 바 있다(The World Bank, 2014). 실제로 영국의 데이터 정책을 살펴보면, 정부가 높은 관심을 갖고 공공 부문의 데이터 제공과 관리 역할을 담당하는 동시에 민간 부문의 데이터 공개 참여와 확대, 범국가적 데이터 역량 강화, 연구를 위한 경우 적극적인 데이터 개방과 활용 지원 등 상당히

적극적이고 주도적인 역할을 담당하는 것을 볼 수 있다. 우리나라에서도 정부를 비롯한 공공기관의 높은 관심과 주도성, 그리고 그에 기반한 국가 차원의 데이터 전략 설계가 주요할 것으로 판단되는 바, 정부 관계자, 정책 결정자들의 데이터 관련 인식과 관심, 데이터 정책 추진을 위한 전문적 지식과 역량, 리더십 제고가 시급하다 하겠다.

② 학계, 기업, 시민단체 등 다양한 이해관계자의 참여와 협력

현재 고안 중인 「국가 데이터 전략」 수립 과정을 살펴보면, 정부가 주도적으로 데이터 관련 법제와 정책을 추진하고 있으나 그 과정에 있어서는 학계, 기업, 시민단체, 일반 국민 등 다양한 이해관계자의 참여와 협력을 추구함을 볼 수 있다. 예를 들어, 「국가 데이터 전략」 수립을 주관하는 DCMS는 정책의 구체적 설계에 앞서 공공 및 민간(기업, 시민단체, 학교와 연구기관 등) 부문 그리고 일반 국민을 대상으로 현황조사(Call for Evidence)를 실시하여 사람, 경제, 정부의 세 차원에서 데이터 관련 인식과 관심, 실태, 욕구 및 현안 등을 파악하였다. 또한 DCMS는 공공부문과 기업, 시민단체 등 250여 곳에서 참가자들을 모아 20회 이상의 원탁 회의도 진행하며 다양한 이해관계자의 참여를 독려하고 그 의견을 생생하게 수렴하고 있다. 따라서 영국의 빅데이터 정책은 정부가 정책 추진을 이끌고 있지만 구체적인 정책 설계는 모든 이해관계자들이 참여하고 함께 협력하여 만들어가는 정책이라 할 수 있다.

이러한 참여적, 협치적 정책 추진의 특징은 열린 정책 설계와 실험을 전문으로 하는 Policy Lab(정책실험실), 데이터 개방과 활용, 그를 통한 혁신을 추구하는 비영리 연구기관 Open Data Institute(오픈 데이터 연구소) 등의 데이터 정책과정 참여와 협력에서도 발견할 수 있다. Policy Lab은 앞서 언급한 국가 데이터 전략 수립을 위한 자료 수집과 분석, 그리고 원탁회의 구성과 운영을 담당하며 정책 설계과정에 참여하고 있고, 오픈 데이터 연구소는 영국이 일찍이 오픈 데이터 정책을 추진하는 가운데 긴밀하게 협력해 왔다.

이러한 참여적, 협치적 추진과정은 빅데이터 정책, 특히 아동·청소년 관련 빅데이터 정책을 모색함에 있어 중요한 시사점을 제공한다. 즉 빅데이터 정책이 데이터를 개방하고 그에 대한 분석과 활용을 촉진함으로써 보다 나은 가치를 창출하고자 하듯이, 빅데이터 정책 수립 과정 또한 모든 이해관계자에게 열려있어 다양한 사람들의 의견이 자유롭게 제기되고 그에 관해 활발하게 토론하며 정책화가 이루어질 수 있어야 하며, 이러한 과정을 통해 궁극적으로 사회적 가치가 창출되어야 한다는 것이다. 오픈 데이터 연구소가 ‘변화 이론(Theory of Change)’에서 개방성과 신뢰성을 강조한 것도 바로 동일한 맥락이라 할 것이다.

③ 데이터 인프라와 역량, 데이터 양과 질을 모두 추구하는 통합적 접근

영국 정부는 데이터 정책을 설계함에 있어 물적 차원이라 할 수 있는 인프라와 인적 차원이라 할 수 있는 역량, 그리고 데이터에 있어서도 양적 차원 곧 데이터량과 속도와 접근성, 그리고 질적 차원 데이터의 질과 윤리 등을 모두 포괄하는 통합적 접근을 추구하고 있다. 빅데이터 정책은 결국 데이터의 저장과 관리, 데이터에 대한 접근성(네트워크) 등의 인프라와 데이터를 분석, 해석, 활용 및 관리할 수 있는 전문적 역량, 데이터가 수집, 축적되어 분석, 활용 및 관리되는 전 과정에 있어서 윤리와 책임성, 그리고 이 모든 것이 가능하기 위한 전제라 할 수 있는 데이터 개방에 대한 사회적, 정책적 합의 등을 모두 포함하는 것으로 어느 한 부분만으로 온전하고 실효성 있는 정책일 수 없기 때문이다.

이는 우리나라에 대하여도 다르지 않을 것이다. 앞서 제언한 국가 차원의 데이터 전략 수립에 있어, 인프라 구축, 데이터 분석 및 활용 전문 인력 양성, 국가 차원의 데이터 분석과 활용 역량 강화, 데이터 기반의 정책과 연구 체계 구축, 책임 있고 윤리적 데이터 수집과 분석, 활용, 관리를 위한 사회적 합의 및 구체적 법제도 마련, 데이터 윤리의식 함양 등이 분절되어 추진되는 것이 아니라 통합적 접근으로 다루어져야 할 것이다.

④ 빅데이터 기반 연구 체계 구축과 지원

영국의 빅데이터 관련 정책 추진에서 주목할 만한 점 또 하나는 바로 빅데이터 기반 정책 연구 체계인 ADR UK(Administrative Data Research UK)이다. ADR UK는 전국의 대학과 연구기관 등과 협력하여 주택과 커뮤니티, 건강과 웰빙, 아동·청소년, 직업 세계, 고령화, 불평등과 사회통합, 기후위기와 지속가능성, 범죄와 사법의 8개 주제 아래 58개 연구 프로젝트 수행을 승인, 지원하고 있다. 이때 영국 정부가 빅데이터 기반 연구 체계가 데이터 기반의 과학적, 전문적 연구를 활발히 할 수 있도록, 연구 목적의 경우 공동 데이터가 보다 적극적으로 개방될 수 있도록 법적 근거를 마련하고 경제사회연구위원회를 통해 재정적 지원을 제공하는 한편 ADR UK의 연구 활동에 대하여 정기적으로 모니터링하고 그 결과를 의회 등에 보고하는 등 법적, 행정적, 재정적 기반을 담당하고 있는 점이 인상적이다.

빅데이터에 기반한 연구를 통해 보다 나은 정책 설계 또는 결정을 이끌 수 있다는 점에서 우리나라에서도 빅데이터 기반 연구 체계 구축이 시급하다. 빅데이터 정책이 공공 데이터를 비롯한 빅데이터 개방을 논하는 것에 그치는 것이 아니라 빅데이터 활용을 통한 혁신을 도모하는 방향으로 나아가야 하며 이때 빅데이터의 주요 활용 중 하나로 빅데이터 기반 연구 활성화가 논의되어야 할 것이다. 영국 정부의 법제적 지원, 재정적 지원 그리고 행정적 관리감독 역할이 좋은 참고가 될 수 있을 것이다.

3. 호주의 아동·청소년 빅데이터 체계 구축 및 활용 사례

1) 호주의 빅데이터 정책

호주의 빅데이터 전략의 개발은 ‘APS ICT 전략 2012-2015’⁶⁰⁾에서 처음으로 제시하였으며, 이는 정부기관에서 보다 향상된 서비스의 제공 및 개발을 지원하도록 각 기관의 데이터 분석능력을 향상시키기 위해 빅데이터에 대한 정부의 전체적인 접근방식의 필요성을 강조하였다. 재정규제완화부(Department of Finance and Deregulation) 내의 호주정부 정보관리사무소(The Australian Government Information Management Office, AGIMO)가 이러한 작업을 주도하고 있다.

2013년 3월 15일 AGIMO는 AGIMO 블로그에 Big Data Strategy - Issues Paper를 발표하였다(Department of Finance and Deregulation (Australia), 2013). 이 전략의 목적은 국민, 산업 및 학계와 빅데이터에 대한 공개적인 대화를 시작하는 것이었다. 이러한 공개적인 협의과정에서 수집한 의견은 빅데이터 전략의 개발에 사용되었다.

이 전략의 개발과 병행하여 호주 국세청은 정부 전체의 데이터분석 우수센터(Data Analytics Centre of Excellence, DACoE)를 출범시켜 데이터분석을 주도하였다. DACoE는 분석을 위한 공통기능 프레임워크를 설정하고, 기술적인 지식, 기능 및 도구를 공유하고, 분석전문가의 양성을 위해 제3의 기관과 협력계약을 구축함으로써 정부 전반에 걸쳐 분석기능을 구축할 예정이다. DACoE는 빅데이터 전략 및 빅데이터 작업그룹의 기타 작업을 통해 수립된 지침 내에서 작동한다.

본 장에서는 호주 연방정부에서 공표한 공공서비스 빅데이터 전략(2013년)과 2019년의 데이터 전략을 요약하였다.

60) Department of Finance and Deregulation, Australian Public Service Information and Communications Technology Strategy 2012-2015, http://agimo.gov.au/ict_strategy_2012_2015/

(1) 호주의 공공서비스 빅데이터 전략

호주의 재정및규제완화부(Department of Finance and Deregulation, DFD)는 빅데이터의 가치는 통찰력을 추출하고, 더 나은 의사결정을 내리는 능력에 있다는 기본개념 하에 향상된 데이터분석 역량을 통하여 사물 및 현상에 대한 이해를 증진하고자 공공서비스 빅데이터 전략계획을 2013년에 발표하였다⁶¹⁾. 빅데이터 전략계획은 2013년 2월에 설립하여 학계, 산업계, 정부기관 등의 다기관 실무그룹(빅데이터 실무그룹)에서 의견을 수렴하여 개발되었다.

호주의 정부기관이 보유한 데이터는 오랫동안 정부 및 국가의 자산으로 인식되어 왔다. 새로운 정보통신기술을 수용하고, 그리고 정부 밖에서 생성되는 정형 및 비정형 데이터가 증가하면서 데이터의 잠재적 성장이 있었다.

빅데이터 분석을 사용하여 서비스 제공을 간소화하고, 혁신을 위한 기회를 창출하고, 새로운 서비스 및 정책 접근방식을 식별할 수 있을 뿐만 아니라 국가 인프라를 유지 관리하고, 정부 운영 전반에 걸쳐 기존 프로그램의 효과적인 제공을 지원할 수 있다. 보건의료분야의 경우, 응급구조요원의 대응시간을 단축하여 보건의료 서비스 제공을 향상할 수 있다.

빅데이터 전략은 서비스 제공 및 정책 개발을 담당하는 호주 정부기관의 고위 관료를 대상으로 한다.

전략계획은 빅데이터에 대한 정의, 자산으로서 데이터의 인식, 빅데이터 개방 및 공유에 따른 개인정보의 보호, 데이터 관리방법 등을 서론에서 요약하였으며, 빅데이터 활용의 기회와 편익, 빅데이터의 비전, 원칙, 그리고 실행계획을 포함하고 있다.

빅데이터와 이를 분석하고 이해하는 데 도움이 되는 도구 및 기법은 정부기관에 기회와 과제를 명확하게 제시한다. 실제로 새로운 개념으로서 빅데이터는 이 개념

61) Department of Finance and Deregulation(Australian), Australian Government Big Data Strategy, 2013(https://toolkit.data.gov.au/images/1/16/Big_Data_Strategy.pdf)

이전의 많은 데이터 관리문제와 서로 교차한다. 빅데이터는 연방기관이 보유한 데이터의 가치문제와 호주 대중에게 이익을 제공하기 위해 이 가치를 실현할 책임은 물론 데이터의 연결, 공유 및 광범위한 액세스 제공에 따른 개인정보보호의 위험을 협상할 필요성을 다시 제기한다.

이 전략은 호주의 정부기관이 운영을 지원하기 위해 빅데이터 사용을 고려할 때 따를 수 있는 경로를 제공하는 것을 목표로 한다. 또한 개인의 프라이버시를 보호하는 동시에 더 나은 서비스 제공 및 정책개발을 통해 기관의 생산성 향상을 지원하는 것을 목표로 한다.

① 빅데이터의 기회와 편익

정부기관은 구조화, 반구조화, 비구조화 등 다양한 형식으로 제공되는 점점 더 많은 데이터를 보유하거나 액세스할 수 있다. 호주 정부기관에서만 증가하는 데이터 생산에 대처하기 위해 2008년부터 2012년까지 93,000TB의 스토리지를 추가로 설치하였다.

빅데이터는 조직에 광범위한 잠재적 기회와 혜택을 제공한다. 빅데이터가 제공할 가치의 규모와 성격은 산업분야에 따라 다르지만, 정부는 빅데이터를 사용하여 상당한 생산성과 혁신을 실현할 수 있을 것으로 예상된다.

일부 빅데이터 프로젝트는 비즈니스 문제에 대한 예상치 못한 통찰력을 제공할 것으로 예상된다. 이는 빅데이터 분석과정이 전통적인 방법과 부분적으로 상이하기 때문이다. 전통적인 과학적 탐구는 가설을 설정하고 데이터 수집을 통해 가설을 검증하는 것으로 진행되지만, 빅데이터 분석은 반대 방향으로 작동할 수 있다. 분석을 통해 결론을 도출할 수 있는 통찰력을 보여주는 기존 데이터의 양이 많다. 그러나 빅데이터 프로젝트에 대한 결과 지향적 접근방식은 여전히 매우 중요하다.

산업계의 경험에 따르면 이러한 예상치 못한 상관관계와 발견은 다른 방법으로는 도달할 수 없었던 혁신적인 솔루션으로 이어질 수 있는 중요한 통찰력을 제공할

수 있다. 이러한 통찰력은 정보와 추세가 발생할 때, 보다 신속하게 대응하고 대응할 수 있는 기회를 제공할 수도 있다.

이에 전략계획에서는 빅데이터의 기회와 편익을 서비스 제공, 정책 개발, 통계, 비즈니스 및 경제적 기회, 빅데이터 분석 기능(skill), 생산성 등의 차원에서 아래와 같이 정리하고 있다.

가. 서비스 제공 차원의 기회와 편익

빅데이터를 이용하는 기관들이 패턴과 상관관계를 활용하여 서비스를 제공하고, 결과를 개선하는 혁신적인 방식으로 빅데이터 분석을 활용할 수 있는 잠재력을 인식하는 것이 중요하다. 이러한 통찰력은 기관이 서비스, 정책 및 프로그램을 더 잘 조정하고 타겟팅하도록 지원함으로써 생산성과 효율성을 높이는 데 도움이 될 수 있다. 개선된 서비스 타겟팅은 기관이 과잉 서비스를 더 잘 관리하고 예방하는 동시에, 서비스를 받을 자격이 있으면서 인식하지 못하는 사람들이 받을 수 있는 서비스를 놓치지 않도록 보장하는 데 도움이 될 것이다.

개선된 서비스 제공은 적절한 건강 및 복지 서비스의 적시 제공, 주요 인프라 관리, 개인화된 사회보장 혜택 제공, 개선된 응급서비스, 정부 및 민간 부문 모두에서 사기 또는 범죄활동 및 오류 감소 등과 같은 다양한 영역을 포괄할 수 있다. 더 많은 공공영역의 정보(public sector information, PSI)가 계속해서 제3자에 의해 재사용 가능하게 됨에 따라 혁신적인 새로운 서비스의 개발이 가능할 것이다.

빅데이터 분석은 의사결정자에게 더 많은 증거를 제공하여 특정 개인 및 커뮤니티의 요구와 관심을 반영하는 서비스를 맞춤형하는 방법을 가능하게 할 수 있다. 소비자의 요구를 반영한 서비스의 개인화는 정부서비스에 더 간단하고 쉽게 액세스할 수 있도록 하며, 과도한 서비스의 방지 또는 사람과 지역사회에 대한 서비스의 더 나은 대응(매칭)을 가능하게 하여 서비스 제공비용 절감에 도움이 될 수 있다.

나. 정책 개발 차원의 기회와 편익

“증거 기반 정책 입안의 성공여부”는 그 기반이 되는 증거의 질에 달려 있다. 정부가 다양한 출처의 풍부한 정보계층에 액세스하고 분석을 수행할 수 있도록 함으로써 더 나은 정부 정책과 결과를 얻을 수 있다.

APS200 Project40의 최종 보고서⁶²⁾는 APS 및 과학기관 내에서 그리고 APS 및 과학기관 간의 지식 관리, 통합 및 공유가 정책을 지원하기 위해 데이터 및 연구 서비스에 대한 액세스 및 사용을 촉진할 수 있다고 언급했다.

빅데이터를 더 잘 분석할 수 있기 때문에 의사결정자는 다양한 정책 옵션을 모델링하고 정책이 구현되기 전에 정책의 결과를 더 정확하게 예측할 수 있으며, 정보를 사용하여 정책개발 프로세스를 알리고 개선할 수 있다. 그런 다음 기관은 이 세분화된(granular) 정보를 사용하여 더 나은 정보에 입각한 보다 대응력 있는 결정을 내리고, 더 짧은 시간에 더 낮은 비용으로 커뮤니티에 원하는 결과를 달성할 수 있다.

의사결정과 관련한 빅데이터 분석의 힘에 대한 증거는 연속적인 데이터 분석에서 거의 “실시간” 통찰력을 제공하는 능력이다. 이것은 종종 “nowcasting”으로 설명된다. Google의 수석 이코노미스트 Hal Varian이 빅데이터의 미래와 현재에 대하여 다음과 같은 의견을 제시하고 있다.

“나는 Nowcasting을 믿는다. 거의 모든 대기업에는 실시간 데이터웨어하우스가 있으며, 정부기관보다 경제에 대한 시의 적절한 데이터가 있다. 향후 10년 동안 우리는 정부가 이러한 민간부문 데이터저장소의 일부를 활용할 수 있도록 하는 공공-민간 파트너십을 보게 될 것이다. 이는 더 나은 정보에 입각한 보다 적극적인 재정 및 통화 정책으로 이어질 것이다.”

62) Innovation, APS200 Project: The Place of Science in Policy Development in the Public Service, 2012 (<http://www.innovation.gov.au/Science/Pages/APS200ProjectScienceinPolicy.aspx>)

빅데이터의 실시간 분석은 (특히 긴급대응이 필요한 경우) 공중보건, 사회서비스 및 환경 위협을 포함한 다른 정책 영역에 대한 단서를 제공할 수 있다. 의사결정자에게 주제에 대한 최신 정보를 제공하는 것도 향후 정책 이니셔티브에 대한 방향을 제공하여 새로운 문제에 대하여 사전적으로 정책대응을 가능하게 한다. 이러한 예측 기능은 정부가 자연재해와 같은 위협이 발생하기 이전에 더 잘 관리하는 데 도움이 될 수 있다.

다. 통계 차원의 기회와 편익

빅데이터는 경제, 사회 및 환경 문제에 대해 정부와 대중에게 알리는 수단으로서 중요하게 기여할 수 있다. 전통적으로 공식적인 통계는 거의 전적으로 정부 프로그램과 설문조사 데이터 수집에 의해 수집된 행정 데이터를 기반으로 하였다. 이러한 데이터 소스는 여전히 중요하지만, 비교적 구조화되지 않은 데이터 소스에서 적절하게 통합하여 수집된 빅데이터의 분석결과와 결합되면 더 관련성이 높고, 통찰력 있고, 시의적절한 통계가 정부에 제공되어 더 나은 정책 및 서비스 제공에 대한 의사결정을 할 수 있다.

라. 비즈니스 및 경제 차원의 기회와 편익

많은 산업계 지지자들은 운영의 최적화, 더 나은 정보에 입각한 의사결정 도구 제공, 재무 및 기타 위협의 관리 및 완화, 새로운 비즈니스 모델의 개발 등 빅데이터 분석이 제공하는 실용적인 비즈니스 기회를 확인하였다. 이러한 모든 기회는 생산성과 혁신의 증가로 이어질 것이다. 빅데이터 분석의 발전은 또한 완전히 새로운 산업에 대한 기회를 창출하고 있다.

산업계 전문가들은 호주의 과학 및 학술 기관에서 나오는 빅데이터 연구개발 결과의 상업화를 인정해야 한다고 강조했다.

광대역, 통신 및 디지털 경제부 (DBCDE)는 최근 디지털 경제를 선도하는 호주의 입지를 확보하기 위한 여러 이니셔티브를 요약한 국가 디지털 경제 전략 (National Digital Economy Strategy)을 수정하여 발표했다. 이 전략은 혁신과 생산성 향상의 촉진자로서 빅데이터와 오픈 데이터의 중요성을 인식한다. 이 전략은 또한 주 정부 수준의 개방 데이터 공간에서 발생하는 작업을 인식하고 지원한다. 예를 들어, 정부의 투명성, 책임성 및 효율성을 지원하는 최근 발표된 NSW 정부의 개방 데이터 정책 초안이 있다.

마. 빅데이터 분석 기능 (Skills) 차원의 기회와 편익

빅데이터 분석은 새로운 ICT 일자리와 심지어 새로운 직업을 창출할 잠재력을 가지고 있다. 많은 관찰자들은 현재 빅데이터 분석 데이터 과학자에게 큰 기술격차가 있음을 지적했다.

정부는 ICT 인력문제를 인식했다. 갱신된 국가 디지털 경제 전략은 기술에 대한 새로운 커리큘럼 개발을 완료하고, 학생들의 ICT 경력을 향상하기 위한 이니셔티브를 제공하였다.

GovHack⁶³⁾은 빅데이터 분석의 핵심인 데이터 매시업, 앱 및 시각화에 대한 기술과 관심의 개발을 더욱 촉진하고 지원한다. GovHack은 팀이 오픈 데이터로 혁신적인 솔루션을 생산할 수 있는 새로운 방법을 찾도록 권장하는 48시간의 경쟁 이벤트이다. 산업, 연구 및 학술 부문은 한동안 빅데이터 분석 프로젝트를 진행해 왔으며, 빅데이터 분석과 관련된 기능(skills), 기술(technologies), 기법(techniques)에 지속적으로 투자하고 있다. 이러한 부문은 공익을 촉진하는 빅데이터 분석에서 통찰력을 제공하기 위해 귀중한 데이터 수집의 핵심 관리자이자 정부와의 잠재적인 파트너로 인식되고 있다.

63) GovHack은 호주 최대의 개방 정부이며, 개방형 데이터 해커톤으로 매년 1000명이 넘는 참가자를 유치한다. 2009년 캔버라에 기반을 둔 소규모 행사로 처음 운영되었으며, 매년 전국 주요 도시에서 동시 행사가 개최되는 전국대회로 빠르게 확장되었다.

정부는 빅데이터 분석 및 관련 분야의 전문지식을 활용하고 공유함으로써 이러한 부문과 협력할 것이다.

정부는 또한 Data Analytics Center of Excellence와 같은 이니셔티브를 통해 기관 전반의 기술을 강화하고 있다. 이 이니셔티브의 목적은 분석기능을 구축하는 동안 기술 지식, 기능 및 도구를 공유하기 위해 정부 및 다양한 분야의 대표자들을 모으는 것이다.

정부는 또한 다수의 빅데이터 시범(pilot) 프로젝트를 시작하여 학습을 늘리는 방법을 모색하고 있다. 이러한 파일럿 프로젝트는 정부의 운영방식을 개선하고 개인에게 실질적인 가치를 제공할 수 있는 빅데이터 분석의 잠재력을 보여주는 데 도움이 될 것이다.

바. 생산성 차원의 기회와 편익

공공부문에서 빅데이터 분석을 효율적으로 사용하는 것은 국제 관할권에서 생산성 향상의 잠재적인 동인으로 확인되었다. 예를 들어, McKinsey Global Institute는 유럽의 공공부문이 행정활동비용을 잠재적으로 15 ~ 20%까지 절감할 수 있다고 추정했다. 이 추정에 따르면 절감액은 약 1,500억 ~ 3천억 유로에 해당한다. McKinsey는 또한 향상된 서비스 및 정부 효율성의 결과로 향후 10년 동안 연간 생산성이 0.5% 포인트까지 증가하는 것으로 확인했다.

영국의 정책교환(UK Policy Exchange) 조직은 빅데이터 도구를 사용하여 사기와 오류를 줄이고 '세금 격차'(실제 징수된 세금과 이론적 부채의 차이)를 해소함에 있어서 더 큰 생산성을 달성할 수 있다고 추정한다. 이러한 감축으로 연간 16억 ~ 330억 파운드를 절약할 수 있다.

② 호주 정부의 빅데이터 비전

호주 정부는 빅데이터 분석을 사용하여 서비스를 개선하고, 새로운 서비스를

제공하며, 더 나은 정책 조언을 제공하는 동시에 개인정보보호를 위한 모범사례를 통합하고 기존의 ICT 투자를 활용하고자 한다. 이에 호주정부는 공공부문에서 효율성, 협업 및 혁신을 추진하기 위해 빅데이터 분석을 사용하는 세계적인 리더가 되는 비전을 수립하였으며, 이러한 비전을 달성하기 위하여 다음과 같은 역량들을 지원하고 있다.

가. 향상된 서비스

서비스의 제공 결과에 대한 더 나은 정보를 제공하고, 이러한 서비스 제공에 대한 미래 모델을 알리고, 현재 서비스 제공 계약에서 격차가 존재하는 위치를 식별한다.

- 정부기관이 서비스를 필요로 하는 사람들에게 더 나은 서비스를 제공할 수 있도록 하여 보다 효율적이고 효과적인 서비스 제공을 허용한다.
- 기업과 지역사회의 개별 요구에 따라 서비스 제공을 맞춤화하여 공공기관이 서비스를 개선할 수 있도록 한다.

나. 새로운 서비스 및 비즈니스 파트너십의 기회

빅데이터 분석은 분석 프로세스에서 파생된 통찰력을 기반으로 하는 새로운 서비스의 개발로 이어질 것으로 예상된다.

- 산업 발전과 빅데이터 분석을 활용하는 도구 및 서비스의 성숙은 공개된 정부 데이터 사용을 기반으로 완전히 새로운 비즈니스 기회와 산업을 창출할 것이다.

다. 개선된 정책 개발

증거 기반의 의사결정을 강화하여 더 나은 정책개발을 지원하고, 정책의 환경 및 그것의 영향에 대한 보다 즉각적인 정보를 제공한다.

라. 프라이버시 보호

빅데이터 분석 프로젝트에 “설계된 개인정보보호”를 통합하고, 개인의 데이터 및 정보에 대한 개인정보보호를 사전에 보장한다.

- 빅데이터 분석 및 “모자이크 효과”로 인한 잠재적인 프라이버시 위험을 해결하는 더 나은 방법론을 채택한다.

마. ICT 기술에 대한 정부의 투자 활용

National Broadband Network와 같은 정보통신기술에 대한 정부의 투자를 활용한다.

- 클라우드 기술과 함께 전국 광대역 네트워크를 사용하여 빅데이터 분석 프로젝트의 진입 비용을 낮춘다⁶⁴⁾.

③ 6가지 빅데이터 원칙 (Big data principles)

호주의 공공서비스 빅데이터 전략에서는 다음의 빅데이터 원칙을 제시하고, 각 기관은 이러한 원칙에 기반하여 빅데이터에 접근할 것을 안내하였다.

가. 원칙 1 : 데이터는 국가 자산이다

정부가 보유한 데이터 세트는 국가의 자산이며, 공익을 위해 사용해야 한다.

- 개방 정부의 선언 및 기타 입법 요건에 따라 이 데이터를 공유하면, 참여 문화가 향상될 것이다.

나. 원칙 2 : 개인정보보호 설계

빅데이터 프로젝트는 “디자인에 의한 개인정보보호”를 통합한다.

64) Department of Finance and Deregulation, Australian Government Cloud Computing Policy, <http://agict.gov.au/policy-guides-procurement/cloud>

- 이는 빅데이터 프로젝트의 전체 수명주기 동안 개인정보보호 및 데이터가 보호됨을 의미한다.
- 모든 데이터 공유는 관련 법률 및 비즈니스 요구사항을 준수한다.

다. 원칙 3 : 데이터의 무결성 및 프로세스의 투명성

빅데이터 분석을 사용하여 향상된 서비스를 제공하는 기관은 빅데이터 프로젝트에 대한 동료 검토 및 공개적인 자문을 수행하도록 권장한다.

- 기관은 새로운 빅데이터 프로젝트에 대해 개인정보 영향평가 (PIA)를 수행하고, 이러한 PIA를 게시하도록 권장한다⁶⁵⁾.
- 빅데이터 분석 프로젝트의 각 당사자는 다음 사항에 대한 책임을 인식하고 준수해야 한다.

라. 원칙 4: 기술, 리소스 및 기능은 공유된다.

데이터 분석에 대한 기술과 전문지식은 적절한 경우 정부기관 및 산업계의 기관 간에 공유된다.

- 데이터 세트 및 분석 모델에 필요한 인프라는 적절하고 가능한 경우 기관 간에 공유된다.
- 빅데이터의 분석 역량은 DACoE와 같은 노력을 통해 정부 차원의 접근방식으로 강화될 것이다.

마. 원칙 5 : 산학 협력

산업체, 연구기관 및 학계 등은 그동안 빅데이터 분석 프로젝트를 진행해 왔으며, 빅데이터 분석과 관련된 기술 및 기법들에 지속적으로 투자한다.

65) Office of the Australian Information Commissioner, Privacy Impact Assessment Guide, <http://www.oaic.gov.au/privacy/privacy-resources/privacy-guides/privacy-impact-assessment-guide>)

- 정부기관은 연구부문을 빅데이터 분석의 통찰력을 제공하는 핵심 파트너이자 중요한 데이터 수집의 주요 생산자 및 관리자로 인식한다.
- 정부기관은 빅데이터 분석에 대해 산업체, 학계, 비정부기관 및 기타 관련 당사자들과 지역 및 국제적으로 협력한다. 이러한 참여는 빅데이터 워킹 그룹, 호주정부의 최고기술책임자 및 정부전체의 DACoE를 통하여 장려한다.
- 이러한 참여는 빅데이터 분석에 대한 민간 및 공공 부문의 경험과 전문성을 활용하고, 이 분야에서 정부기관의 지식과 기술을 향상시킬 것이다.

바. 원칙 6 : 오픈 데이터 강화

기관들이 아웃소싱 및 혁신 장려를 목적으로 정보를 공개하도록 권장함에 따라 오픈 데이터는 계속해서 향상될 것이다.

정부기관은 개방형 공공영역의 정보(PSI) 원칙에 따라 빅데이터 분석 프로젝트에 접근할 것이다. 이러한 원칙은 PSI가 커뮤니티 액세스 및 사용을 위해 사용 가능하고 검색 가능한 국가자원이라는 정부2.0(Gov 2.0)의 전제에 기초한다. 정부기관은 이러한 고려사항과 상충되는 법적 의무 (예 : 보안 및 개인정보보호 관련)의 균형을 맞춰야 한다.

정부의 빅데이터 프로젝트의 결과는 공개되고, 분석과정에서 사용되거나 생성된 데이터 세트는 대중의 고려와 소비를 위해 공개 라이선스 하에 data.gov.au에 공개된다.

빅데이터 분석은 Gov 2.0 구현을 기반으로 한다. data.gov.au 포털은 개선될 것이며, 검색 및 사용 가능한 정부의 정보를 위한 중앙 데이터 저장소 역할을 계속할 것이다.

④ 6가지 행동계획

호주 공공서비스 빅데이터 전략계획에서는 아래와 같이 6가지 행동계획을 제시하였다.

가. 조치 1 : 빅데이터 우수사례지침의 개발 [2014년 3월까지]

빅데이터 워킹그룹은 DACoE와 협력하여 빅데이터 분석에서 정부기관의 역량을 향상시키는 것을 목표로 하는 더 나은 실행지침을 개발할 것이다. 이 지침은 다음을 수행한다.

- 기관이 빅데이터 분석이 개선된 서비스 제공 및 더 나은 정책개발을 지원할 수 있는 위치를 식별하도록 지원하는 조언을 포함한다.
- 빅데이터 분석 이니셔티브에 필요한 거버넌스 준비를 식별한다.
- 기관이 가치 있는 데이터 세트를 식별하도록 지원한다.
- 정부의 제3자 데이터 세트 사용 및 제3자의 정부 데이터 사용에 대해 조언한다.
- 의도적으로 프라이버시를 증진한다.
- 프라이버시 영향평가 (PIA)를 촉진하고, 동료 검토 및 품질보증 프로세스를 명확히 한다.
- 클라우드 컴퓨팅 사용에 관한 정책 및 지침에 대한 참조를 포함한다⁶⁶⁾.

이 지침에는 관련 기관의 기존의 조언도 포함된다. 예를 들어, 이 지침은 국가통계청(National Statistical Service, NSS)에서 만든 통계 연구 목적을 위한 연방 데이터와 관련된 데이터 통합 원칙⁶⁷⁾을 참조한다. 이 지침은 NSS에서 개발한 통계적 공간 프레임워크 (SSF)⁶⁸⁾를 참조할 것이다. 이 프레임워크는 공간적으로 사용 가능한 정보의 접근성과 유용성을 개선하기 위한 관점에서 사회 경제적 및 위치 데이터 통합에 대한 일반적인 접근방식을 제공한다. 또한 기관이 데이터 및 정보를 익명화하는 데 도움이 되는 리소스를 포함하여 OAIC에서 생성한 문서를 참조할 것이다⁶⁹⁾. 이 지침을 준비할 때 산업계와 학계의 의견을 구할 것이다.

66) 클라우드 컴퓨팅에 대한 자세한 정보 및 지침:

(<http://agimo.gov.au/policy-guides-procurement/cloud/>)

67) 국가 통계청. 통계 및 연구 목적을 위한 연방 데이터를 포함하는 데이터 통합을 위한 높은 수준의 원칙 (<http://www.nss.gov.au/nss/home.nsf/NSS/7AFDD165E21F34FDCA2577E400195826?opendocument>)

68) 국가 통계청. 통계적 공간 프레임워크

(<http://www.nss.gov.au/nss/home.NSF/pages/Statistical%20Spatial%20Framework>)

이 지침은 또한 빅데이터 분석 프로젝트를 수행할 때 위험 평가 및 보안관리에 대한 조언을 제공한다.

나. 조치 2 : 빅데이터 분석의 장벽 식별 및 보고 [2014년 7월까지]

빅데이터 워킹그룹은 DACoE와 협력하여 정부에서 빅데이터의 효과적인 사용에 대한 장벽을 식별할 것이다. 이러한 장벽에는 기술, 정책, 입법 기술, 자원, 조직 및 문화 등의 장벽이 포함된다. 모든 장벽을 해결할 수는 없지만, 이러한 장벽을 자세히 설명하고, 가능한 완화 및 개선 전략과 조치를 고려하는 보고서가 생성된다.

다. 조치 3 : 빅데이터 분석의 기술 및 경험 향상 [2014년 7월까지]

빅데이터 워킹그룹은 DACoE와 협력하여 빅데이터 분석을 활용하는 기존 프로젝트와 선정된 정부기관이 주도할 새로운 빅데이터 프로젝트 착수를 포함한 여러 빅데이터 파일럿 프로젝트를 식별하고 지원한다. 이러한 파일럿 프로젝트는 학습, 혁신 및 협업을 촉진함으로써 빅데이터 관련 기술의 개발을 향상시킬 것이다.

또한 빅데이터 워킹그룹은 DACoE와 협력하여 교육 커리큘럼 향상을 목표로 하는 모든 이니셔티브에서 ICT의 광범위한 기술과 함께 고려될 빅데이터 분석을 위한 다양한 특정 기술을 옹호할 것이다. 예를 들어, 이러한 기술에는 정보 및 커뮤니케이션 기능, 정보학 및 통계, 수학, 사회 경제학, 비즈니스, 언어학 및 영향 평가기술 등이 포함될 수 있다.

라. 조치 4 : 책임 있는 데이터 분석 가이드 개발 [2014년 7월까지]

빅데이터 워킹그룹은 책임 있는 데이터 분석에 대한 가이드를 개발하기 위해 DACoE와 협력할 것이다. 이 가이드는 빅데이터 프로젝트의 거버넌스에 초점을 맞추고, 개인정보보호와 관련하여 OAIC의 권장사항 및 지침을 통합한다. 이 가이

69) 호주 정보 커미셔너 사무실. 익명화 리소스 2013년 5월
(<http://www.oaic.gov.au/privacy/privacy-engaging-with-you/previous-privacy-consultations/de-identification-resources-may-2013/>)

드에는 국가통계청(NSS), 교차 포트폴리오 데이터 통합 감독위원회, 사무국의 역할에 대한 기관정보 등도 포함된다⁷⁰⁾.

이 가이드는 통계 및 연구 목적을 위한 연방 데이터를 포함하는 데이터 통합을 위한 NSS에서 생성한 높은 수준의 원칙⁷¹⁾을 통합할 것이며, 여기에는 기관이 연방기관이 보유한 데이터의 통합을 포함하는 빅데이터 프로젝트를 개발할 때 기관이 사무국과 상호작용하는 방법과 시기가 포함된다. 이 가이드는 또한 이러한 프로젝트를 지원하기 위한 투명한 검토 프로세스의 잠재력을 조사할 것이다.

마. 조치 5 : 정보자산 레지스터 개발 [진행 중]

빅데이터 워킹그룹은 DACoE와 협력하여 기관별 정보자산 등록부 개발을 지원하는 기관에 대한 지침을 생성한다. 이러한 정보자산 레지스터는 재사용 가능한 데이터 세트에 대한 기관 간의 가시성을 지원한다. 이 조치는 기관 전반에 걸친 정부 2.0(Gov 2.0) 구현을 기반으로 하며, Commonwealth 기관이 보유한 데이터를 더 잘 관리하고, data.gov.au에 공개되는 데이터 세트의 수를 늘리는 데 도움이 될 것이다. 이 지침은 PSI⁷²⁾ 게시 지침 및 data.gov.au 이니셔티브 관련 작업을 포함한 기존 문서를 활용한다.

바. 조치 6 : 빅데이터 분석기술의 발전을 적극적으로 모니터링. [계속]

AGIMO의 지원을 받는 빅데이터 워킹그룹의 구성원은 빅데이터 분석기술의 발전을 적극적으로 모니터링하고 산업, 연구 및 학술 전문가를 요청하여 워킹그룹에 업데이트를 제공할 것이다.

70) 국가 통계청. Commonwealth 데이터와 관련된 통계 데이터 통합.

(www.nss.gov.au/dataintegration)

71) National Statistical Service, 통계 및 연구 목적을 위한 연방 데이터를 포함하는 데이터 통합을 위한 상위 수준 원칙,

<http://www.nss.gov.au/nss/home.NSF/pages/High+Level+Principles+for+Data+Integration+-+Content?OpenDocument>

72) 호주 정부 웹 가이드, 공공부문 정보 게시,

<http://webguide.gov.au/web-2-0/publishing-public-sector-information/>

(2) 호주의 데이터 전략

호주의 재정및규제개혁부(Department of Finance and Deregulation, DIIS)는 2013년에 호주 공공서비스 빅데이터 전략계획(The Australian Public Service Big Data Strategy)을 발표하면서 빅데이터 기반의 국가경제전략을 지속적으로 추진하여 왔다⁷³⁾.

2018년에는 호주의 산업혁신과학부(Department of Industry, Innovation and Science, DIIS)가 “자료전략 2018-20 (Data Strategy 2018-20)⁷⁴⁾”을 발표하여 신뢰하는 데이터 및 분석을 의사결정에 사용하여 가치를 부여하고, 자료 중심의 국가 경쟁력 향상 및 경제 혁신을 지향하고 있다. 이 계획은 이전의 공공 빅데이터 개방 및 활용 관련한 각 부처의 계획에 연계하여 자료 기반의 국가경쟁력 향상을 추구하고 있다. 이 계획의 비전, 주요 전략, 기존의 계획들과의 연계에 대한 내용을 요약하면 아래와 같다.

비전

정책 조언, 프로그램 제공, 서비스 구현 및 의사결정에서 양질의 데이터 및 증거를 생성하여 공유하고 적용하여 비즈니스 경쟁력을 높이고 경제 혁신을 촉진함으로써 호주인의 웰빙에 기여한다.

우리는 우리가 하는 모든 일에서 데이터의 힘을 활용하는 것을 목표로 한다. 우리는 데이터를 중심으로 데이터를 평가하고, 데이터의 분석기능을 활성화하고, 국민들이 데이터에 잘 액세스하고 사용하는 데 대한 장벽을 제거한다.

향후 2년 동안 우리는 데이터 자산의 완전한 가치를 실현하기 위해 체계적이고 문화적인 변화를 만들 것이다. 우리는 호주 기업과 시민의 성과를 개선하기 위한

73) Department of Finance and Deregulation, The Australian Government Big Data Strategy (2013) (https://toolkit.data.gov.au/images/1/16/Big_Data_Strategy.pdf)

74) Department of Industry, Innovation and Science. Data Strategy 2018-20. Australian Government, 2018 (<https://www.industry.gov.au/sites/default/files/2019-01/data-strategy-2018-20.pdf>)

연구, 정책 조언, 프로그램 및 서비스를 뒷받침하는 증거 기반을 계속 구축할 것이다.

선별되고 신뢰할 수 있으며 검증된 데이터를 모든 사람이 쉽게 사용할 수 있도록 보장하고, 부서 내 데이터 및 분석기능에 투자하고, 일상적으로 데이터와 통찰력을 소중히 여기는 문화를 구축하여 다음을 수행한다.

- 국민의 데이터 활용 능력, 참여 및 권한 부여 향상
- 증거 기반의 즉각적이고 신속한 결정 및 선택 추진
- 정책 설계 및 조언, 프로그램 및 서비스 제공 개선
- 기업, 지역 및 개인에 대한 투자 영향 이해

전략적 제휴 (strategic alignment)

호주의 데이터 전략은 보다 나은 데이터 거버넌스 및 관리, 포트폴리오 및 부문 간 데이터 공유 및 사용, 윤리적인 데이터 사용, 개인정보 보호 및 보안을 보장하기 위한 호주 정부의 목표와 일치한다. 이것은 호주 정부의 각 부처에서 발표하여 추진하는 다수의 정부 개혁, 데이터 정책 및 지침 등에 연계하여 추진한다(아래 참조).

- 데이터 가용성 및 사용 문의 및 호주 정부의 대응 (생산성위원회)
- 1988년 개인정보보호법의 파트 IIIC에 따른 통지 가능한 데이터 위반 (Notifiable Data Breaches) 체계
- 호주의 2030 보고서 및 호주 정부의 대응 (호주 혁신 및 과학부)
- 호주정부 빅데이터 전략 (2013); 공공 데이터 정책 성명서 (2015); 디지털 연속성 정책 2020
- 열린 정부 (2015년)의 공약 및 제 2차 열린 정부 파트너십 국가 행동계획
- Bizlab, 기타 혁신 연구소, 호주 정부의 행동경제팀 등 설립
- Data61을 통해 고가치 및 확장 가능한 프로젝트를 제공하는 오픈 데이터 이니셔티브용 플랫폼

- 다기관 데이터 통합 프로젝트, 경제데이터 분석 네트워크 (Economic Data Analysis Network, EDAN), 비즈니스 종단자료 분석 환경 등을 포함하여 호주 데이터 통합 파트너십이 주도하는 국가 데이터 통합 프로젝트
- 호주와 뉴질랜드 10개 정부부처의 국가차원의 공간 데이터를 포함하는 재단 공간 데이터 프레임워크
- 호주 공공부문 프레임워크 및 호주 공공부문위원회 데이터 리터러시 가이드의 데이터 기능 및 역량

호주의 데이터 비전 실현을 위하여 사람들에게 데이터 기능 및 역량을 제공하고 권한을 부여, 데이터 분석을 중시하는 문화 조성, 품질을 보장하여 가치 있는 데이터의 용이한 사용, 분석 및 시각화를 사용하여 가치 중심 의사결정 등 4가지 전략을 수립하였다.

① 호주의 2018-2020 데이터 전략 구상

가. 데이터 관리 및 데이터 문화 조성

산업혁신과학부(The Department of Industry, Innovation and Science)는 호주의 성장과 일자리 창출을 주도한다. 우리는 비즈니스, 과학, 혁신 및 투자를 지원함으로써 호주의 경쟁력을 높이고 경제 변화를 촉진한다. 이를 달성하기 위해 데이터 및 분석의 힘을 활용하여 영향력 있는 정책과 관행을 제공하는 데 필요한 강력한 증거를 확보하고 있다.

이 데이터 전략(Data Strategy 2018-20)은 현재까지의 데이터 성과를 기반으로 하며 양질의 데이터를 생성, 공유 및 채택할 수 있는 역량에 더 많이 투자할 것을 약속한다. 데이터를 업무의 핵심 요소로 만들고 데이터 사용을 문화적 규범으로 포함하기 위한 비전과 계획을 설명한다. 이 전략은 과제를 해결하고 생산성을

높이려면 매일 데이터를 소중히 여기고 사용해야 한다는 점을 강조한다.

더 많은 데이터 중심이 되려면 리더십과 우리가 생각하고 행동하고 행동하는 방식을 변화시켜야 한다. 우리는 다양한 아이디어, 강점 및 도구를 활용하고, 열린 대화를 나누고, 널리 듣고, 협력하며, 정보에 입각한 결정을 내릴 것이다. 우리는 데이터를 사용하여 가장 필요한 시기에 도움이 필요한 기업과 개인을 지원하는 정책을 목표로 삼을 것이다. 우리는 또한 우리의 데이터에 표현된 사람들의 신뢰를 얻어야 한다. 데이터 전략은 정부 전체 데이터 의제에 대한 우리의 지속적인 리더십과 영향력을 나타낸다.

데이터의 힘을 활용하고 진행상황을 검토하며 영향을 측정하는 데 도움이 되는 지원계획 및 이니셔티브를 개발할 것이다.

2020년까지 사용 가능한 데이터를 최대한 활용하기 위해 다음을 수행한다.

- 국민들이 데이터 활용 기능과 역량을 갖도록 권한을 부여한다.
- 민첩하고, 대담하며 참여도가 높은 의사결정자의 공유문화를 조성하여 의사결정자가 정보에 입각한 행동, 방어 가능한 정책, 프로그램 및 서비스 결정 등을 하도록 한다.
- 부서가 보다 고객-중심, 성과-지향 및 결과-중심으로 전환하도록 자산으로서 데이터의 가치를 매긴다.
- 신뢰할 수 있고 선별된 데이터를 광범위하고 안전하게 사용하도록 한다.
- 데이터를 통합하여 기존 데이터의 사용을 확대한다.
- 기존의 데이터 거버넌스 및 관리 관행을 확장한다.
- 데이터 프라이버시 보호 및 소셜 라이선스 관리를 보장한다.
- 데이터 인프라 및 분석역량을 구축한다.
- 세련된 통찰력을 창출할 수 있는 기회를 발견한다.
- 정부차원의 데이터 이니셔티브가 정책에 영향을 주도하도록 유도한다.

나. 데이터 역량 및 성과

데이터 관리 전략 2016-18을 통하여 지금까지의 성과를 아래와 같이 요약하고 있다.

- 데이터 자산을 검색하고 액세스 가능
- 신뢰의 기반으로서의 양질의 데이터 모델 구축
- 통합 데이터 분석
- 데이터의 통제 및 위험관리
- 데이터 실행 커뮤니티(communities of practice) 운영
- 가치 있는 데이터 세트 인증
- 데이터 활용능력 향상
- 정부간 데이터 자산
- 데이터에 대한 신뢰할 수 있는 조연자 제공

다. 데이터 중심의 조직으로 변화하기 위한 청사진

정책 조언, 프로그램 제공, 서비스 실행 및 결정 등을 함에 있어서 양질의 데이터와 증거를 생성, 공유 및 적용하여 비즈니스 경쟁력을 높이고, 경제 변화를 촉진함으로써 호주인의 웰빙에 기여하였다. 이에 데이터 중심의 조직으로 변화할 것을 강조하며 청사진을 제시하였다.

표 IV-9 변화를 위한 청사진 (1/2)

구분	변화를 위한 세부내용		
데이터 처리	<ul style="list-style-type: none"> • 각 부처의 데이터 집단을 위한 인식, 지식 및 수용을 점진적으로 높여 왔다. • 데이터 사용자 및 데이터 관리자를 위한 데이터 교육을 실행하였다. • 데이터는 부서를 통해 모든 수준에서 완전히 평가되지 않다. • 데이터에 정통한 더 많은 사람들이 높은 수준의 데이터를 잘 사용하고, 사용요구를 리드하도록 기초적인 데이터 역량이 필요하다. 	<ul style="list-style-type: none"> • 데이터의 사용자, 분석가, 관리자 및 임원에 대한 교육으로 데이터 활용능력을 향상하자. • 직원이 MyPlan을 통해 데이터 학습여정을 관리하도록 장려한다. • 데이터 가치, 원칙 및 행동을 홍보하여 데이터 문화를 구축하자. 	<ul style="list-style-type: none"> • 사람들은 기본적인 데이터 기능을 가지고 있고, 협업하고, 데이터를 사용하여 문제를 해결하고, 데이터를 우선하는 마음자세와 작업 방식을 가지고 있다. • 관리자는 데이터를 사용을 권장하자. • 산업혁신과학부는 협업, 다양성, 실험 등으로 데이터 혁신문화를 구축한다.
	<ul style="list-style-type: none"> • 데이터 관리 전략 2016-18은 우수한 데이터 거버넌스 및 관리를 지원하였다. • 데이터가 중앙 집중화 되고, 검색 가능하다. 	<ul style="list-style-type: none"> • DataHub를 통해 보다 신뢰성 있는 선별된 데이터에 쉽게 접근하게 하자. • 데이터 공유를 향상하자. • data.gov.au에 더 많은 데이터를 게시하자. • 사전 예방적으로 데이터를 관리하자. • 분석 프레임워크 및 지침을 개발하자. • 보고를 넘어 예측적 분석으로 이동하자. • 데이터의 신뢰 및 소셜 라이선스를 구축하자. 	<ul style="list-style-type: none"> • 풍부한 데이터 및 분석으로 일상적인 엔터프라이즈 기능에 정보를 제공한다. • 예측 및 추세 예측에서 의사결정의 품질 및 속도를 향상한다. • 데이터 무결성을 통해 의사결정의 신뢰성을 향상한다. • 더 많은 데이터를 공개적으로 사용할 수 있다.

표 IV-10 변화를 위한 청사진 (2/2)

구분	변화를 위한 세부내용		
데이터 기술	<ul style="list-style-type: none"> 점차 양질의 데이터세트, 스마트 도구 및 리소스를 증가시켜 왔다. 데이터 아키텍처는 사용 사례와 디지털 기술에 의해 주도된다. 데이터를 잘 사용하려면 더 많은 도구와 더 나은 IT 시스템이 여전히 필요하다. 	<ul style="list-style-type: none"> 데이터와 디지털 도구 및 기술을 활성화하여 결과를 최적화하자. Trial Data Science Lab, 모델 및 인프라 (인공지능, 예측 및 규범분석 포함)를 구축하자. 위치 인텔리전스를 위한 공간적으로 데이터 세트를 활성화 하자. 	<ul style="list-style-type: none"> 분석준비가 완료된 통합 데이터 고급 분석 지원 직원은 업무에 맞는 다양한 분석 도구 중에서 선택 가능 올바른 도구를 사용하여 데이터 스토리텔링을 지원하도록 데이터 시각화
데이터 정책	<ul style="list-style-type: none"> 데이터 거버넌스 및 관리 관행이 성숙해지고 있다 데이터 침해 정책이 마련되어 있다. 데이터 수집이 항상 일관적이거나 시의적절한 것은 아니다. 	<ul style="list-style-type: none"> 엔터프라이즈 데이터 모델, 원칙 기반 품질 및 표준 접근방식을 아웃 효과적이고 안전한 데이터 실행을 위한 중요한 지침으로 데이터 거버넌스 및 가드 레일 설정 	<ul style="list-style-type: none"> 공통 데이터 정의 사용 사용된 데이터 및 데이터에 대한 신뢰가 증가함 데이터 침해가 최소화됨 좋은 데이터 관행 준수

* 출처: Department of Industry, Innovation and Science. Data Strategy 2018-20. Australian Government, 2018.

라. 데이터 전략 설계 원칙

가치를 창출하기 위해 전략은 데이터 문화에 대한 기존의 장벽을 해결하고, 우리가 이용할 수 있는 기회를 활용하기 위한 조치를 취해야 한다. 이를 위하여 해결해야 할 문제에 대하여 도전하고, 기회를 이용하기 위한 전략설계의 원칙을 설정하였다.

가) 도전 (해결해야 할 문제)

- 사람 : 데이터 사용에 대한 국민의 불신이나 개인정보보호 고려사항은 데이터 공유를 방해할 수 있다.
- 프로세스 : 데이터가 프로그램 전반에 걸쳐 적시에 또는 일관된 방식으로 수집되지 않거나 의사결정과 단절되어 고객의 상황 이해 또는 스토리에 영향을 방해할 수 있다.
- 기술 : 기술의 제한점으로 인해 직원이 도구에 액세스하여 통찰력을 얻을 수 없는 경우가 발생한다.
- 문화 : 사람들은 변화에 개방적이지 않고 디지털 혁신의 속도를 따라갈 수 없다.

나) 기회

- 리더십 : 우리는 성장, 생산성, 경쟁력 및 혁신을 추진하기 위한 정부 의제를 중심으로 한다.
- 증거 : 우리는 추측이나 직감이 아닌 데이터와 경험적 증거로 결정을 내린다.
- 신뢰 : 우리는 양질의 데이터를 사용할 수 있도록 직원, 정부, 기업 및 시민과 신뢰, 신뢰 및 책임을 공유한다.
- 다양성 : 우리는 다양한 관점, 관점 및 도구를 제공하고 수용한다.

- 복잡성 : 우리는 새로운 작업방식을 수용하고, 혁신적이고, 민첩하며, 창의적이며, 학습 및 확장을 지원하고, 위험에 대비하는 긍정적인 문화를 만든다.
- 협업 : 우리는 데이터 이니셔티브에 대해 파트너 및 시민과 협력하고, 공통의 목표와 비전을 식별하고, 집단적 이익과 영향을 달성하는 것을 목표로 한다.
- 리터러시 (문해력) : 우리는 사람들이 데이터를 읽고, 데이터로 작업 및 분석하고, 데이터 스토리를 전달하고, 데이터에 대해 논쟁할 수 있도록 재교육 및 지원한다.

마. 전략적 조정 또는 전략의 연계 (Strategic alignment)

호주 데이터 전략의 성공을 위하여 국가데이터를 보유하고 있는 기관 및 데이터 간 전략적 조정 및 연계가 절대적으로 필요하다. 정부 데이터 자산의 사용 및 가치 극대화, 정부의 사업보조금 관리 간소화, 디지털 데이터 기반의 비즈니스 모델로 전환 등 3가지 측면에서 그 방안을 정리하였다(표 IV-11 참조).

표 IV-11 전략적 조정(연계)의 주요 측면들

주요 측면들	세부 내용
정부 데이터 자산의 사용 및 가치 극대화	<p>데이터 전략은 더 나은 데이터 거버넌스 및 관리, 포트폴리오 및 부문간 데이터 공유 및 사용, 윤리적 데이터 사용, 개인정보보호 및 보안을 보장하기 위한 호주 정부의 목표와 일치한다. 이는 다음을 포함한 정부 전반의 여러 개혁, 데이터 정책 및 방향에 의해 주도되고 있다.</p> <p>생산성위원회의 데이터 가용성 및 사용 문 의와 호주 정부의 대응</p> <ol style="list-style-type: none"> 1. 1988년 개인정보보호법 파트 III C에 따른 통지 가능한 데이터 위반 (Notifiable Data Breaches, NDB) 체계 2. Innovation and Science Australia의 호주 2030 보고서 및 호주 정부의 대응 3. 호주 정부 빅데이터 전략 (2013); 공공 데이터 정책 성명 (2015); 디지털 연속성 정책 2020 4. 열린 정부 (2015) 공약 및 제 2차 열린 정부 파트너십 국가 행동 계획 5. Bizlab의 설립(기타 혁신연구소 및 호주 정부의 행동 경제팀) 6. Data61을 통해 높은 가치의 확장 가능한 프로젝트를 제공하는 공개데이터 추진용 플랫폼

주요 측면들	세부 내용
	7. 다기관 데이터 통합 프로젝트, 경제 데이터 분석 네트워크 (EDAN) 및 비즈니스 종단분석 데이터 환경을 포함하여 호주 데이터 통합 파트너십이 주도하는 국가 데이터 통합 프로젝트 8. Foundation Spatial Data Framework : 호주 및 뉴질랜드 10개 정부부처의 국가 수준의 공간 데이터를 포함 9. 호주 공공부문 프레임워크의 데이터 기능 및 역량, 그리고 호주 공공부문위원회 데이터의 리터러시 가이드
정부의 사업보조금 관리 간소화	우리는 Business Grants Hub를 통해 정부보조금 관리를 간소화하는 데 기여한다. 이제 기업들이 한 곳에서 호주정부의 지원 프로그램에 쉽게 접근할 수 있다. 우리의 데이터와 디지털 이니셔티브는 이를 위해 매우 중요하다. 당사의 정책 파트너는 신뢰할 수 있는 고품질 프로그램 관리, 성과보고 및 결과 측정을 제공하는 부서에 의존한다.
디지털 데이터 기반의 비즈니스 모델로 전환	데이터 전략은 부서의 디지털 전략 2017-20과 연계한다. 이들은 데이터와 디지털 솔루션을 함께 활용하여 비즈니스 프로세스의 자동화 향상 및 데이터 기반 의사결정 지원을 포함하여 부서의 전략적 방향에 기여함으로써 운영모형을 변화시킬 것이다.

* 출처: Department of Industry, Innovation and Science. Data Strategy 2018-20. Australian Government, 2018.

② 호주의 4가지 데이터 전략 (2018-20년)

네 가지 전략을 수립하여 데이터 비전을 지원한다. 전략의 핵심은 직원이 일상 업무에서 데이터와 분석을 사용하는 데 필요한 기술과 능력을 부여하는 것이다. 리더와 직원이 함께 모여 데이터와 분석을 소중히 여기는 조직문화를 조성할 것이다. 마지막으로, 신뢰할 수 있는 데이터와 분석도구를 사용하여 우리가 일하는 방식을 변화시키고 선도적인 호주 공공서비스(Australia Public Service, APS) 기관으로 자리매김하는 것을 목표로 한다. 4가지 데이터 전략은 아래와 같다.

가. 전략 1 : (사람과 능력) 장비, 데이터 기술, 능력을 가진 인력 등의 역량 강화 (사람들에게 데이터 기술 및 기능을 제공하고 권한을 부여한다)

사람들에게 데이터 기술 및 기능을 제공하고 권한을 부여한다. 학습 및 개발 프로그램을 통해 직원은 효과적인 지식근로자가 되기 위한 기초 데이터 기능에

대한 교육을 받는다. 직원은 데이터 식별, 큐레이팅, 분석 및 시각화에 특화된 기능을 제공할 수 있는 데이터 관리자, 데이터 관리자 및 데이터 과학자의 지원을 받는다.

표 IV-12 데이터의 기능 및 역량

<p><u>데이터 관리</u></p> <ul style="list-style-type: none"> - 데이터 관리자 (stewards) - 데이터 보호자 (custodians) - 데이터 자산관리 - 데이터 수집 - 데이터 라이선스 및 조달 - 데이터 품질 - 데이터 프라이버시 및 보안 	<p><u>고급 데이터 분석</u></p> <ul style="list-style-type: none"> - 재현 가능한 연구 및 분석 - 통계적 추론 - 회귀 모델링 - data wrangling(원자료를 다른 형태로 전환 또는 mapping) - 고급 데이터 시각화 - 고급 코딩 및 프로그래밍 언어(예 : R, Python)
<p><u>데이터 사용</u></p> <ul style="list-style-type: none"> - 모든 직원 - 데이터 활용 능력 - 데이터 시각화 - 증거 해석 	<p><u>데이터 과학</u></p> <ul style="list-style-type: none"> - 예측 분석 - 기계학습 - 러닝 - 무작위 대조시험 - 최적화
<p><u>데이터 분석</u></p> <ul style="list-style-type: none"> - 섹션당 최소한 한 명의 데이터 분석가 - 설명 및 탐색적 분석 - 데이터 정리 - 데이터 시각화 - Excel 및 Power BI 	<p><u>전략적 데이터</u></p> <ul style="list-style-type: none"> - 의사결정자 및 임원 - 데이터 및 증거 중심의 사고방식 - 전략적 데이터 리더 - 데이터 팀 구축 - 데이터 거버넌스의 안전성, 신뢰할 수 있는 데이터 사용

* 출처: Department of Industry, Innovation and Science. Data Strategy 2018-20. Australian Government, 2018.

나. 전략 2 : (문화) 데이터 분석방법에 대한 가치를 인정하는 (데이터 분석을 중시하는) 문화 조성

데이터 분석을 중시하는 문화를 조성한다. 직원이 데이터에 대해 생각하고 데이터 스토리에 대해 호기심이 많고 호기심이 많은 공유 데이터 비전을 홍보한다.

우리는 매일 데이터 사용에 대한 일련의 가치, 태도, 신념 및 행동을 공유한다. 함께, 우리는 학습, 질문 및 호기심 많은 조직을 구축할 것이다.

다. 전략 3 : 양질의 가치 있는 데이터의 사용성 제고 (품질과 귀중한 데이터를 쉽게 사용할 수 있도록 한다)

품질과 귀중한 데이터를 쉽게 사용할 수 있도록 한다. 전략적 자산으로서 데이터는 전체 수명주기 동안 통제되고 관리된다(획득 시점에서 저장, 공유 및 사용에 이르기까지). 우수한 데이터 거버넌스 관행이 포함된 풍부한 정보는 최종 사용자가 효과적으로 검색하고 신뢰할 수 있는 저장소를 만든다.

라. 전략 4 : (분석 및 시각화) 분석방법 및 정보의 시각화를 통하여 가치 중심의 의사결정 지원

분석 및 시각화를 사용하여 가치 중심의 의사결정을 알린다. 정교한 데이터 분석기술을 사용하여 혁신을 주도하고, 의사결정의 품질과 속도를 높일 것이다. 분석 및 시민 데이터 과학자는 데이터를 매력적인 스토리를 전달하는 통찰력 있는 시각화로 전환한다. 우리의 자원과 투자의 성능, 생산성 및 영향이 증가할 것이며, 우리의 도전과제를 해결하고, 위험을 완화하는 능력이 향상될 것이다.

2) 호주의 청소년 빅데이터 현황

호주의 청소년 관련 공공 데이터는 정부의 공공 데이터 개방 전략에 따라 공개되고 있으며, 호주 정부에서 운영하고 있는 공공 데이터 개방 및 공유 웹사이트(www.data.gov.au)에서 검색 및 활용이 가능하다.

청소년 관련 데이터 세트를 검색한 결과, 검색어 'youth'로 검색한 결과 1,278개, 'children'으로 검색한 결과 761개의 데이터 세트가 검색되었다.

(1) 청소년 관련 빅데이터 개방 현황

데이터를 개방하는 기관들은 보건부, 통계청, 호주보건복지연구원(AIHW), 청소년 관련 각종 정부기관 등으로 구성되어 있다.

호주 청소년의 데이터 거버넌스, 프레임워크, 구축 및 활용을 주도하는 기관은 호주보건복지연구원(AIHW)이라 할 수 있다. 그 밖에 아동 및 청소년 관련 연구기관 또는 주정부에서 데이터의 연계 및 연구를 추진하고 있다.

데이터 셀의 수는 총 2,039개이며, 자료유형은 API, CSV(Excel), pdf, Excel, HTML 기반의 Dashboard (Word, pdf, Excel) 등으로 다양하며, 데이터들은 원자료, 통계 등으로 제공되고 있다(표 IV-13 참조).

표 IV-13 호주의 청소년 개방 데이터 set의 수 및 유형 (data.gov.au)

검색어	데이터 셀(수)	데이터 유형
youth	1,278	API, csv(Excel), pdf, Excel, HTML(dashboard: Word, pdf, Excel)
children	761	

* 출처: data.gov.au (2020년 8월 15일 검색 결과).

(2) 기관의 빅데이터 개방을 지원하는 도구(Toolkit)

호주 정부는 개방데이터의 수준을 개선하기 위하여 각 기관에서 유용하고 실질적인 계획을 수립 가능하도록 하는 toolkit을 제공하고 있다⁷⁵⁾. 툴킷의 주요 내용은 다음과 같다.

- 개방데이터 자체평가표 (open data self-assessment)
- 개방데이터 생성 전략 (creating open data strategies)
- 사례 연구 (case studies)

75) Planning, (<https://toolkit.data.gov.au/Planning.html>)

① 개방데이터 자체평가표

조직 또는 팀의 역량을 고려하여 현재까지 팀의 개방 데이터 성과와 개선 기회를 평가할 수 있는 평가표를 제공한다(표 IV-14 참조).

표 IV-14 개방데이터 자체 평가표

평가기준	기준별 수준 (1, 2, 3, 4 중에서 해당하는 수준을 선택, 번호가 높을수록 수준이 높음)
개방 데이터 인식	<ol style="list-style-type: none"> 1. 주요 고위 직원은 개방 데이터가 무엇이며, 조직에서 어떻게 사용될 수 있는지 이해한다. 2. 모든 직원은 개방 데이터의 관련성을 이해하고, 이를 조직의 핵심 비즈니스에 사용한다. 3. 직원은 조직의 데이터 분석과 그것이 정부, 경제 및 사회에 어떻게 도움이 될 수 있는지에 액세스할 수 있다. 4. 정부 데이터의 구현 및 사용에 대한 직원의 이해를 정기적으로 검토한다.
개방 데이터 관리	<ol style="list-style-type: none"> 1. 개별 직원은 문서화된 데이터 관리 프로세스를 사용한다. 2. 데이터 관리 정책 및 절차는 조직 전체에 알려져 있고 사용된다. 3. 조직의 데이터 관리 절차를 준수하는지 확인하기 위해 비즈니스 프로세스를 정기적으로 검토한다. 4. 조직 전체의 데이터 관리는 지속적인 외부 벤치마킹을 포함하여 측정기준으로 추적된다.
개방 데이터 창조	<ol style="list-style-type: none"> 1. 주요 직원은 데이터를 공개 형식으로 제공하기 위한 문서화된 지침을 알고 있다. 2. 조직의 모든 직원은 일상 업무의 일부로 개방형 데이터 지침을 따른다. 3. 개방형 데이터 지침은 전체 정부 접근방식과 일치한다. 4. 오픈 데이터 지침의 지속적인 개선을 보장하기 위해 공식적이고 문서화된 프로세스가 마련되어 있다.
개방 데이터 배포	<ol style="list-style-type: none"> 1. 내부 및 외부의 이해관계자는 데이터 포털, 클리어링하우스 및 피어 네트워크를 포함하여 조직의 오픈 데이터를 찾는 위치를 알고 있다. 2. 데이터 액세스 및 사용에 대한 공식적인 조치는 조직간에 존재한다. 직원은 동료 네트워크에 참여할 의무가 있다. 3. 사용자 그룹 및 이해관계자 참조 패널은 정기적으로 참여하여 데이터 자산에 대한 피드백을 받고, 조직의 시스템에 액세스 가능성을 확인하고, 개방형 데이터 전략을 정부 전반의 비즈니스 동인에 맞게 조정한다. 4. 조직의 모든 부서는 데이터의 검색 가능성, 접근성 및 적시성을 개선하기 위해 지속적으로 노력한다.
역량 향상	<ol style="list-style-type: none"> 1. 모든 정보관리 프로젝트는 조직에서 인식하고 체계적으로 관리한다. 비즈니스 프로세스에 대한 링크가 포함된 데이터 관리표준이 구현되고 교육을 위한 리소스가 조직의 관련 부분에 할당된다. 2. 모든 데이터 관리기관 프로세스는 중앙지점을 통해 관리되며 정의된 사용자 요구사항과 연결된다. 내부 벤치마킹은 조직 전체의 데이터 관리 및 릴리스 실태(관행)를 비교한다. 3. 데이터 공개 비율은 정량화되어 기관 전체에서 비교된다. 프로세스와 교육은 보편적으로 이용 가능하고 정량화된다. 4. 프로세스 효율성과 다양한 관리 프로세스의 정량화된 측정을 기반으로 지속적인 개선 프로세스가 마련되어 있다.

평가기준	기준별 수준
	(1, 2, 3, 4 중에서 해당하는 수준을 선택, 번호가 높을수록 수준이 높음)
정부간 협력	1. IT 및 기술 팀은 데이터를 사용하여 조직의 다른 사람과 피어 네트워크를 통해 다른 조직에 서비스를 제공한다. 2. 팀 간의 협력에 대한 공식적인 허가와 다른 조직과의 협력을 위한 공식화된 계약이 있다. 3. 공식적으로 데이터를 제공하기 위한 성능 및 검토 프로세스는 내부 및 외부에서 이루어진다. 4. 서비스 계약이 벤치마킹되고, 협업정보 사용에 대한 결합계약이 적용된다.

* 출처: Planning, (<https://toolkit.data.gov.au/Planning.html>)

② 데이터 개방 전략의 생성

기관의 데이터 개방 전략은 개방 데이터에 대한 정부의 약속이 개별 조직 내에서 어떻게 적용되는지를 식별하는 실용적인 문서이다. 조직은 자신의 데이터 개방 전략을 사용하여 아래의 내용을 식별할 수 있다.

- 개방 데이터에 대한 가치
- 개방 데이터 약속 (이용 가능 및 가능하지 않은 데이터)에 대한 조치 방법
(예 : 개인정보 보호 또는 국가보안을 위해 데이터가 보류될 수 있음)
- 계약자 및 데이터 사용자와 같은 공개 데이터 이해관계자와 상호작용하는 방법
- 이용 가능한 데이터에 대한 지시 순서 및 시간표

자신의 조직을 위한 데이터 개방 전략을 수립하는 경우 아래의 ‘사례구조’와 개방 데이터에 대한 자체 평가를 통해 현재의 진행수준에 따라 조직에 가장 적합한 전략적 데이터 개방 목표를 식별할 수 있다. 공개하여 게시할 데이터세트를 선택하는 페이지는 조직 데이터의 게시 우선순위를 정하는 데 유용할 수 있다.

Queensland 주정부는 자신의 데이터 개방 전략을 다른 주정부들이 학습할 수 있도록 공개하였다⁷⁶⁾. 주요 내용은 원칙, 데이터 개방의 중요성, 전략계획, 개방한 데이터, 개방 일정계획 등으로 구성되어 있다.

76) Department of the Premier and Cabinet Open Data Strategy 2018-2020, <https://www.premiers.qld.gov.au/publications/categories/plans/open-data-strategy.aspx>

사례의 구조

- 소개
 - 전략의 목적 및 대상
 - 배경
 - 지금까지 열린 데이터 성과
 - 오픈 데이터 프로젝트의 범위
 - 데이터 정의
- 전략
 - 원칙
 - 목표 및 목표
 - 보유한 데이터 유형
 - 데이터 평가
 - 릴리스 전략
 - 거버넌스
- 출시용 데이터
- 미래 오픈 데이터 프로그램

③ 개방할 데이터세트의 선택

개방 데이터는 데이터 접근을 개방하는 것이며, 정부기관들의 데이터 개방 유형은 다음과 같이 3가지가 있다.

가. 프로그램의 공간 데이터, 에너지 등급 데이터, 범죄 통계, 관리 데이터 등과 같은 일반적인 활동으로 비즈니스에서 생성된 원시 데이터 : 이 데이터는 종종 데이터베이스에 저장되며 주로 비즈니스 응용 프로그램에 사용된다.

나. 처리된 데이터 : 연례 보고서의 테이블, FOI 로그, 기관의 기능 및 운영을 위해 생성된 기타 데이터와 같은 프로세스에서 발생하는 새로운 데이터. 이는 공개 액세스에 적합한 원시 데이터 세트의 전체보기 일수도 있다.

다. 시스템 데이터 : 웹 분석, 프로젝트 관리, 액세스 로그 및 기타 시스템과 같은 다른 프로세스에서 자동으로 생성되는 데이터.

조직 전체에서 서로 다른 데이터를 식별한다는 것은 기존 데이터 팀에서 벗어나 다른 데이터 세트를 살펴보고 이를 활용하여 서비스, 정책 및 효율성을 개선하는 방법을 의미한다.

④ 공개자료의 우선순위 결정방법

조직에서 공개할 데이터를 찾고 있다면, 다음을 통해 널리 사용되거나 쉽게 개방된 데이터 세트를 식별할 수 있다.

- 공통 요청에 대한 FOI, 의회 및 외부 헬프 데스크 요청 분석. 이를 통해 개별 요청에 동일한 데이터를 제공하고 게시할 대중적인 데이터 세트를 식별하는 데 소요되는 시간과 리소스를 줄일 수 있다.

- 기관이 이미 빠른 정보를 얻기 위해 데이터 형식 또는 PDF 형식으로 게시한 정보를 검토한다. data.gov.au에 데이터를 게시하면 접근성, 재사용, 검색 가능성 및 기관에서 데이터를 재사용할 수 있는 능력이 향상된다. 예를 들어, 연례 보고서, 예산, 보조금, 관리 데이터 또는 필수보고의 표가 모두 유용하다.

- 기관에 어떤 데이터 세트가 있는지 평가하고 공개적으로 사용할 수 있는 경우 더 큰 경제적, 투명성 또는 정책상의 이점을 제공하는 상위 10 개를 식별한다.

- 새로운 서비스 또는 응용 프로그램을 제공하기 위해 데이터가 필요한 위치를 식별하고 서비스 제공 비용을 절감하고 데이터를 기반으로 외부 서비스를 활성화하기 위해 데이터 피드를 공개적으로 제공하는 것을 고려하여야 한다.

- 모든 새로운 시스템에서 생성된 데이터와 처음부터 데이터를 가장 잘 재사용하고 적절하게 노출시키는 방법을 고려하여야 한다.

데이터 개방에 대한 기관의 우선순위는 광범위한 개방 데이터 전략의 일부로 식별되어야 한다.

⑤ 출판용 소싱 데이터

공공부문의 정보관리를 다루는 현행 정책의 핵심 주제는 정부의 정보를 국가자원으로 관리해야 하며, 합당한 이유가 없는 한 재사용 가능해야 한다는 것이다. 이를 달성하기 위해 조직은 다음을 수행할 수 있다.

공개할 데이터셋이 공개데이터 형식으로 출시하기에 적합한 정보의 포함여부를 고려하여야 한다. 많은 정부보고서에는 재무, 경제, 사회 또는 규제 활동이나 동향에 대한 기관 데이터가 포함되어 있지만, 재사용에 적합하지 않은 형식으로 데이터를 제공한다(예 : 재사용 가능한 스프레드시트 형식이 아닌 PDF 파일로 스프레드시트 게시). 이 문제를 해결하려면 보고서와 함께 기본 원시 데이터를 재사용 가능한 개방형 데이터 형식으로 게시하여야 한다.

Department of Human Services (DHS)의 Service Locator 도구를 사용하면 가장 가까운 Centrelink, Medicare 또는 Child Support 위치를 찾을 수 있다. DHS는 data.gov.au에 원시 서비스 로케이터 지리 공간 데이터를 공개하였다. 이를 통해 대중은 데이터를 사용하여 자체 앱을 구축할 수 있다.

정부가 data.gov.au의 AusTender 웹사이트에서 과거 정부계약 데이터를 공개하였다.

⑥ 데이터 공개 및 사용 사례

아래의 링크 모음은 data.gov.au 및 기타 개방형 데이터 리소스가 현재 사용되는 방식의 폭과 가치를 보여준다. data.gov.au의 정부 데이터 사용 사례는 data.gov.au의 데이터를 사용하는 방법을 이해하는 중요한 자원이다.

■ 사례1: 국가지도

National Map은 정부기관의 호주 공간데이터에 대한 지도를 기반한 웹사이트이며, 다음을 위해 설계되었다.

■ 사례 2: Trove

Trove는 도서관, 박물관, 아카이브 및 기타 연구기관의 콘텐츠를 제공하고, 탐색 및 구축할 수 있는 도구도 제공한다. Trove는 전체 텍스트(full text) 디지털 리소스의 증가하는 저장소이지만, 사용자가 구축할 수 있는 메타 데이터 집합, 서비스 집합 및 그 자체로 커뮤니티이기도 하다.

■ 사례 3: 에너지 평가 도구

호주 연방, 주 및 뉴질랜드 정부의 공동 이니셔티브는 매년 호주와 뉴질랜드에서 판매되는 7백만 개 이상의 주요 가전제품에 대한 에너지 효율 데이터 세트를 생성하였다. 산업부는 이 데이터를 사용하여 가정에서 에너지 효율적인 가전제품을 선택하는 데 도움이 되는 모바일 앱을 만들었다. 이 데이터 세트는 Jonathon Moffatt 및 Wai Liu가 GovHack 2014를 위해 만든 에너지 계산기와 같은 개인 프로젝트를 위해 구축되었다.

⑦ 빅데이터 : 호주의 공공서비스 빅데이터 전략

APS ICT 전략에 명시된 바와 같이 재무부는 개인의 프라이버시를 침해하지 않고, 빅데이터가 제공하는 기회를 활용하기 위해 정부가 취하고 있는 조치를 설정하는 빅데이터 전략을 개발하였다.

ICT 거버넌스위원회가 승인한 이 전략은 서비스를 향상시키고 새로운 서비스를 제공하며, 더 나은 정책조언을 제공하는 동시에 개인정보보호 모범사례와 기존 ICT 투자 활용효과를 통합하기 위해 빅데이터에 대한 정부 전체의 (WofG) 접근방식을 제공한다.

⑧ 빅데이터를 위한 호주 공공서비스 모범사례 지침

빅데이터를 위한 APS Better Practice Guide는 호주 정부기관의 빅데이터 분석의 수용을 지원하는 것을 목표로 한다. 이 지침은 빅데이터를 수용하고 사용하

기 위한 주요 고려사항에 대해 기관들에게 조언을 제공한다. 또한 정부가 개인의 프라이버시 권리와 정보보안을 계속 보호하면서 기관이 데이터 자산을 더 잘 활용할 수 있도록 지원하는 것을 목표로 한다.

⑨ 연방정부의 데이터 분석 우수 센터

WofG Data Analytics Center of Excellence (CoE)는 호주 국세청에서 정부 전반에 걸쳐 분석기능을 구축하기 위한 공간으로 설립되었다. CoE의 목적은 분석을 위한 공통기능 프레임워크를 활성화하고, 기술 지식, 기능 및 도구를 공유할 수 있는 기회를 제공하는 것이다. CoE는 또한 분석전문가의 개발을 지원하기 위해 고등교육기관과의 협력계약을 구축하는 데 도움이 된다. CoE는 <http://analyticsspace.net.au/>에서 데이터 분석에 대한 정보를 제공하는 웹 사이트를 개발했다.

3) 호주의 청소년 빅데이터 적용 사례⁷⁷⁾

호주에서 아동의 웰빙 정책의 지원, 모니터링, 평가 등을 위한 빅데이터는 주로 기존의 각각의 목적으로 수집 및 보고되는 조사자료 및 행정자료를 연계(linkage)하는 것을 중심으로 추진되고 있다. 호주 아동의 빅데이터 적용사례는 국가단위 및 (준)주별 데이터의 연결로 구분할 수 있다.

데이터 연결(데이터 일치, 데이터 통합 또는 레코드 일치라고도 함)은 사용자가 여러 데이터베이스의 정보를 결합하는 동시에 개인정보를 보호하여 단일 소스에서 가능한 것보다 훨씬 더 강력한 스토리를 전달할 수 있는 방법이다. 데이터 연결은 최근 몇 년 동안 데이터 소스와 데이터가 나타내는 인구에 대한 이해를 높이기 위한 필수 도구로 자리 매김하고 있다.

77) Australian Institute of Health and Welfare. Scoping enhanced measurement of child wellbeing in Australia, 2019.

결합된 데이터에 대한 새로운 분석을 허용하는 새로운 데이터 세트를 생성하기 위해 데이터 연결은 항상 엄격한 보안 및 개인정보보호 프로토콜에 따라 적절한 데이터 거버넌스 프레임워크 내에서 수행된다는 점에 유의해야 한다. 데이터 연결은 법률, 데이터 공유계약 및 권한부여, 윤리인증과 같은 엄격한 데이터 거버넌스 계약의 매개변수 내에서만 수행된다. 예를 들어 AIHW는 공인된 Commonwealth Data Integration Authority로 AIHW가 프로젝트 거버넌스, 기능, 데이터 관리, 개인정보보호 및 기밀유지를 다루는 엄격한 기준을 충족했음을 의미한다. 연결 프로젝트는 AIHW 윤리위원회 및 모든 관련 데이터 관리자, 그리고 많은 경우 기타 관련 윤리위원회의 승인을 받아야 한다. 이러한 데이터 거버넌스 문제는 AIHW 데이터 거버넌스 프레임워크에 자세히 설명되어 있다.

데이터 연결은 여러 가지 방식으로 중요한 역할을 할 수 있다.

- 경로분석 : 둘 이상의 데이터 소스를 연결하면 아동이 주요 발달단계에서 어떻게 전환하는지, 서비스와 상호작용하고, 삶의 여러 단계에서 결과가 어떻게 나타나는지에 대한 이해를 향상시킬 수 있다. 또한 서비스 제공 전반에 걸쳐 개입이 가장 효과적인 지점을 식별하는 데 도움이 될 수 있다. 이를 통해 장애아동, 재택보호아동 (OOHC) 등 영역 또는 그룹의 패턴을 비교하고 서비스 또는 정책의 측면을 평가할 수 있다.

- 연구 및 평가 : 데이터 연결을 사용하여 결과와 기타 (위험 및 보호) 요인 간의 관계를 조사할 수도 있다. 결과분석은 어떤 서비스 또는 개입이 더 효과적일 가능성이 더 높은지에 대한 통찰력을 제공하고, 모니터링을 위한 지표 (특히 예측)의 선택 및 개발에 정보를 제공할 수 있으며, 미래 고객을 위해 보다 개별화된 서비스 제공을 지원할 수 있는 예측도구도 알려준다.

- 방법 및 데이터 품질 개선 : 두 개 이상의 데이터 세트를 연결하여 비용 효율적인 방식으로 데이터 품질 또는 방법을 개선할 수도 있다. 예를 들어, 신뢰할 수 있는 원주민 식별이 있는 데이터 세트를 식별이 상대적으로 불확실할 수 있는

다른 데이터 세트에 연결하면 전체 분석에서 원주민 데이터의 품질과 적용범위를 개선할 수 있다. 데이터 연결은 또한 장애인 상태에 대한 데이터가 데이터 소스 중 하나에서만 제공될 수 있는 장애인과 같은 선택된 인구집단과 관련된 데이터를 통합하는 데 사용할 수 있다.

이 절에서는 아동의 웰빙 또는 아동과 관련된 서비스 제공 시스템의 측정과 관련된 국가 및 주의 데이터 연계 개발 연구의 사례를 정리하였다. 남호주, 서호주, 뉴사우스 웨일즈 등 여러 주(洲)에서는 데이터 연결을 활용하여 기존 아동의 웰빙 데이터 세트를 강화하고 있다. 이러한 연계를 통해 연구자들은 복잡한 연구질문을 도출하고, 아동의 건강 및 웰빙 경로를 조사하고, 긍정적인 아동발달과 관련된 위험 및 보호 요인을 결정할 수 있다. 이 프로젝트는 아동의 건강 및 발달, 임신 및 출산, 모성건강, 원주민 및 토레스 해협 섬 주민 건강, 장애, 자폐증, 구강건강, 입원, 아동 보호 및 OOHHC, 정신건강, 부상 및 자아 등과 관련된 광범위한 데이터를 연결한다.

각 프로젝트에서 전일제 교육의 첫 해에 호주 아동의 발달상태를 측정하기 위해 고안된 국가 발전 측정치인 AEDC (Australian Early Development Census)의 데이터가 정부의 행정데이터와 연결하였다. AEDC의 포함은 아동의 조기발달 취약성으로 이어지는 경로연구를 촉진하고, 조기개입을 위한 정책 및 프로그램 개발을 알리는 데 도움이 된다. 또한 각 연계 프로젝트 간에는 많은 유사점이 있다.

(1) 국가차원의 데이터 연결

국가 차원에서 데이터 연결은 다양한 건강 및 웰빙 데이터 통합에 점점 더 많이 사용되고 있다. 이 접근방식은 더 많은 사람 중심의 데이터와 정책 및 프로그램 개발 및 평가에 대한 다분야 접근방식을 더 잘 지원하는 데이터를 생성한다.

국가 데이터의 연결은 더 나은 결과 또는 더 나쁜 결과를 초래하는 것으로 알려진 주요 예측요소 식별 및 일상적인 모니터링을 위한 추가 정보를 제공하여 예측을 위한 증거를 강화한다.

(2) 주별 데이터 연결

남호주, 서호주, 빅토리아주 등에서 아동 및 청소년의 데이터 연결을 통하여 아동의 건강 및 웰빙 모니터링 및 향상을 지원하는 연구가 실시되었다.

① 남호주주 유아기 데이터 프로젝트 (SA ECDP) 및 서호주주 텔레톤 어린이 연구소의 발달경로 프로젝트 (WA DPP)

남호주 유아기 데이터 프로젝트 (South Australian Early Childhood Data Project, SA ECDP) 및 서호주 텔레톤 어린이 연구소의 발달경로 프로젝트 (Western Australian Telethon Kids Institute's Developmental Pathways Project, WA DPP)는 아동 및 청소년 발달 결과와 관련된 아동, 가족 및 지역사회 수준의 위험 및 보호 요인을 조사하는 것을 목표로 한다. 이 프로젝트는 조기 개입이 성인기의 사회적 및 건강 불이익을 줄일 수 있는가를 식별하는 데 초점을 맞추고 아동 중심 정책 및 프로그램 개발 및 평가를 알리는 증거 기반을 개발하는 것을 목표로 한다.

Telethon 어린이 연구소(Telethon Institute for Child Health Research)에서 수행한 아동의 발달경로 프로젝트는 서호주 아동과 청소년의 건강과 복지, 교육, 장애, 아동 학대 및 방치, 청소년 비행 결과 등의 경로를 조사하기 위한 것이며, 다학제적이고 포괄적인 접근방식을 취한 획기적인 프로젝트이다.

② 뉴사우스웨일스주의 아동발달 연구 (NSW-CDS)

뉴사우스웨일스주의 아동발달 연구(New South Wales Childhood Development Study, NSW-CDS)는 디자인과 목표가 비슷하지만 아동기의 정신건강 및 웰빙과 관련된 발달경로에 더 초점을 맞추고 있다. NSW-CDS는 2009년 개학 첫 해에 87,000명의 어린이 집단을 대상으로 종단자료를 연결한 연구이다. 이 프로젝트는 행정자료(AEDC)를 활용하여 단면평가하고, 정신건강조사 자

료와 연결하여 호주 아동 및 청소년의 정신건강과 웰빙을 개선하기 위한 정책 및 지역사회 활동을 알리는 증거기반을 도출하였다. NSW-CDS는 또한 아동과 부모 모두의 건강, 정신건강, 아동범죄 참여, 교육 및 아동보호 접촉과 관련된 다세대 연결데이터를 사용한다.

이 프로젝트의 연결된 데이터 세트를 활용하는 연구는 조기발달 취약성(AEDC에 의해 측정됨)과 관련된 여러 위험요소(아동의 부모의 정신질환 및 범죄행위, 부모의 흡연 및 직업을 포함한 주산기 특성, 아동 만성 질환 및 비만, 감염으로 인한 입원, 질병 노출, 학대 및 아동보호 참여, 임신연령 및 모성 연령 등)를 확인하였다. 또한 이러한 결과는 포괄적으로 연결된 데이터 세트를 사용하여 식별할 수 있는 건강 및 웰빙의 궤적의 범위를 강조한다.

③ 빅토리아주의 빅토리아 세대 (Gen V)

Victoria주의 머독 어린이 연구소(Murdoch Children's Research Institute)는 주요 파트너와 협력하여 현재 '빅토리아 세대 (Gen V)'를 개발하고 있다. 이 프로젝트는 참가자의 동의하에 평생 건강 및 교육을 포함한 다양한 기준에 수집한 데이터들을 활용하고 있으며, 어린이의 건강, 발달 및 웰빙과 관련된 연구질문을 지원하는 것을 목표로 한다.

4) 우리나라에 대한 시사점

호주의 아동 및 청소년을 위한 빅데이터 정책, 전략계획, 연구 등을 고찰한 바, 국가 차원에서 공공서비스를 위한 빅데이터 전략계획의 수립 및 추진은 재정 및 규제완화부 및 산업혁신과학부에서 주도하고 있다. 이는 우리나라와 유사하다고 할 수 있다.

아동 및 청소년 빅데이터의 구축, 분석, 활용은 호주보건복지연구원(AIHW)에서 주로 담당하고, 그것의 활용을 주도하고 있는 것으로 파악되었다. 호주의 공공 서비스 향상 및 혁신을 위한 빅데이터 전략 및 사례의 고찰을 바탕으로 다음과 같이 시사점을 도출해 보았다.

첫째, 아동청소년에 대한 빅데이터는 인터넷, 소셜미디어 등 사이버공간의 소셜 데이터를 수집하여 분석 및 활용을 강조하지 않고 있으며, 기존에 여러 목적으로 수집 또는 보고되어온 데이터 세트(행정데이터, 패널데이터, 조사데이터 등이며, 이를 데이터 자산 또는 데이터 원천으로 명명함)를 연계하여 아동의 웰빙 정책 수립 지원, 모니터링, 평가 등에 활용하고 있다. 우리나라도 이러한 것을 참고하여 소셜빅데이터 및 기존의 데이터 세트를 종합적으로 연계하여 분석함으로써 청소년 및 아동의 웰빙 및 발달경로에 대한 연구 및 감시, 추적 등을 실시할 수 있을 것이다.

둘째, 청소년 및 아동의 발달 및 웰빙을 위한 데이터 연계는 국가수준, 주 수준 등 크고 작은 규모의 지리적 특성별로 추진하여 유용한 정보를 획득할 수 있을 것이다. 특히, 지리적으로 다차원의 데이터 세트를 구축하여 아동 및 청소년 웰빙 및 발달을 위한 정책의 수립 지원, 모니터링, 평가 추진에 활용할 수 있을 것이다.

셋째, 아동 및 청소년이 바람직한 성장 및 웰빙을 위하여 국가차원에서 데이터 거너번스 및 프레임워크를 구축하고, 전략계획을 수립하여 빅데이터 연계 방안, 빅데이터 연계에 의한 정책연구, 실행을 위한 분석 및 연구를 촉진하고, 아동 및 청소년 빅데이터 활용사례의 공유에 의한 활성화 등을 리드하는 전담기관을 마련하여야 할 것이다.

넷째, 데이터의 원천별로 청소년 및 아동의 데이터를 구분하고(소셜 데이터, 행정/실적 데이터, 조사데이터), 또한 아동 및 청소년의 영역별 데이터 세트의 구성 및 활용 방안을 아동 및 청소년 빅데이터 전담기구에서 마련하여야 할 것이다.

4. 외국의 아동·청소년 빅데이터 체계 구축·활용 사례 소결 및 시사점

미국국토안보부에서 조사한 2019년부터 2022년까지의 분석결과를 보면, 빅데이터 시장은 매년 17.5%씩 성장하며 국가안보와 공공 안전을 위해 가장 많이 사용될 것으로 전망하고 있으며, 빅데이터를 가장 많이 사용할 것으로 예상되는 상위 국가 중에 미국은 1위, 한국은 11위로 제시되고 있다(USA Homeland Security, 2018; 한국정보화진흥원, 2019, 재인용). 기본적으로 빅데이터는 정형 데이터뿐만 아니라 일상에서 만들어지는 비정형 데이터까지 포함하여, 국민 전체의 삶의 질 향상을 위해 존재한다. 하지만 특히 아동·청소년 분야의 빅데이터는 4차 산업혁명을 바라보는 현 상황에 비추어 국가발전의 주요한 인력자원과 관련한 정책 영역이라는 점에서 연구와 정책개발 등 보다 적극적인 활용 방안이 요구된다.

따라서 아동·청소년 데이터의 생산(production)과 제공(provision), 활용도 제고(promotion), 정보보호(protection) 및 그에 따른 아동·청소년분야의 준비(preparation)와 관련한 청소년 빅데이터 체계구축은 시급하면서도 중차대한 국가발전 의제(agenda)로 논의되어야 할 것이다. 이와 관련하여 호주와 영국 그리고 미국 등의 해외사례를 살펴본 바, 우리에게 주는 시사점을 6가지로 정리, 요약할 수 있다.

첫째, 빅데이터 일반 뿐만 아니라 청소년 빅데이터 체계구축을 위한 정부주도의 노력이 필요하다. 먼저 호주는 아동 및 청소년을 위한 빅데이터 정책, 전략계획, 연구 등을 위하여 국가 차원에서 공공서비스를 위한 빅데이터 전략계획을 수립하고 추진하는 것으로 확인하였다. 영국의 경우도 일찍부터 데이터의 가치에 주목하며 정부 주도의 데이터 정책을 추진해 왔다. 2000년 정보공개법 제정을 시작으로 2009년 「스마트 정부 구현을 위한 실행계획서」, 2012년 「오픈 데이터 백서」, 2014년 「오픈 데이터 전략 2014-2016」 등을 발표하며 공공 데이터 개방 정책을 펼치며 정부 운영의 효율성과 투명성, 궁극적으로 책무성을 제고하고자

하였다. 이후에는 데이터 분석과 활용에 집중하며 2013년 범국가적 「데이터 역량 강화 전략」, 2017년 「디지털 전략」 등을 추진하였고 현재는 「국가 데이터 전략」을 고안하는 중이다. 이처럼 영국 정부는 (빅)데이터를 미래 영국을 이끌어갈 주요한 성장동력으로 여기며, 관련 법제를 마련하고 정책을 추진하는데 매우 적극적인 모습이다.

이러한 정부의 높은 관심은 빅데이터 관련 법적, 행정적 준비를 촉진할 뿐 아니라 학계와 기업, 시민단체 등 민간 부문의 참여와 협력을 보다 강력하게 견인할 수 있기에 중요하다. 세계은행은 보고서에 따르면 빅데이터 시대에 방대한 공공 데이터를 보유하고 있는 정부는 데이터 공급자로서 뿐만 아니라 다른 공공기관, 민간기업, 시민단체 등의 데이터 개방을 위한 선도자이자 촉진자로서의 역할도 담당할 수 있다며 그 중요성을 강조한 바 있다(The World Bank, 2014). 실제로 영국의 데이터 정책을 살펴보면, 정부가 높은 관심을 갖고 공공 부문의 데이터 제공과 관리 역할을 담당하는 동시에 민간 부문의 데이터 공개 참여와 확대, 범국가적 데이터 역량 강화, 연구를 위한 경우 적극적인 데이터 개방과 활용 지원 등 상당히 적극적이고 주도적인 역할을 담당하는 것을 볼 수 있다.

우리나라에서도 정부를 비롯한 공공기관의 높은 관심과 주도성, 그리고 그에 기반한 국가 차원의 데이터 전략 설계가 주요할 것으로 판단되는 바, 정부 관계자, 정책 결정자들의 데이터 관련 인식과 관심, 데이터 정책 추진을 위한 전문적 지식과 역량, 리더십 제고가 시급하다 하겠다. 또한 영국 정부는 빅데이터 기반 연구 체계가 데이터 기반의 과학적, 전문적 연구를 활발히 할 수 있도록, 연구 목적의 경우 공동 데이터가 보다 적극적으로 개방될 수 있도록 법적 근거를 마련하고 경제사회연구위원회를 통해 재정적 지원을 제공하는 한편 ADR UK의 연구 활동에 대하여 정기적으로 모니터링하고 그 결과를 의회 등에 보고하는 등 법적, 행정적, 재정적 기반을 담당하고 있는 점이 인상적이다. 구체적으로 빅데이터 기반 정책 연구 체계인 ADR UK(Administrative Data Research UK)는 전국의 대학

과 연구기관 등과 협력하여 주택과 커뮤니티, 건강과 웰빙, 아동·청소년, 직업 세계, 고령화, 불평등과 사회통합, 기후위기와 지속가능성, 범죄와 사법의 8개 주제 아래 58개 연구 프로젝트 수행을 승인, 지원하고 있다. 이처럼 영국 정부는 빅데이터 연구를 위하여 법적 지원, 재정적 지원 그리고 행정적 관리감독 역할을 생산적으로 수행하고 있다.

둘째, 빅데이터 구축과 활용을 위한 거버넌스가 요구된다. 영국의 경우처럼 학계, 기업, 시민단체, 일반 국민 등 다양한 이해관계자의 참여와 협력이 필요하다. 예를 들어, 「국가 데이터 전략」 수립을 주관하는 영국의 DCMS는 정책의 구체적 설계에 앞서 공공 및 민간(기업, 시민단체, 학교와 연구기관 등) 부문 그리고 일반 국민을 대상으로 현황조사(Call for Evidence)를 실시하여 사람, 경제, 정부의 세 차원에서 데이터 관련 인식과 관심, 실태, 욕구 및 현안 등을 파악하였다. 또한 DCMS는 공공부문과 기업, 시민단체 등 250여 곳에서 참가자들을 모아 20회 이상의 원탁회의를 진행하며 다양한 이해관계자의 참여를 독려하고 그 의견을 생생하게 수렴하고 있다.

따라서 영국의 빅데이터 정책은 정부가 정책 추진을 이끌고 있지만 구체적인 정책 설계는 모든 이해관계자들이 참여하고 함께 협력하여 만들어가는 정책이라 할 수 있다. 이러한 참여적, 협치적 정책 추진의 특징은 열린 정책 설계와 실험을 전문으로 하는 Policy Lab(정책실험실), 데이터 개방과 활용, 그를 통한 혁신을 추구하는 비영리 연구기관 Open Data Institute(오픈 데이터 연구소) 등의 데이터 정책과정 참여와 협력에서도 발견할 수 있다. Policy Lab은 앞서 언급한 국가 데이터 전략 수립을 위한 자료 수집과 분석, 그리고 원탁회의 구성과 운영을 담당하며 정책 설계과정에 참여하고 있고, 오픈 데이터 연구소는 영국이 일찍이 오픈 데이터 정책을 추진하는 가운데 긴밀하게 협력해 왔다. 이러한 참여적, 협치적 추진과정은 빅데이터 정책, 특히 아동·청소년 관련 빅데이터 정책을 모색함에 있어 중요한 시사점을 제공한다. 즉 빅데이터 정책이 데이터를 개방하고 그에

대한 분석과 활용을 촉진함으로써 보다 나은 가치를 창출하고자 하듯이, 빅데이터 정책 수립 과정 또한 모든 이해관계자에게 열려있어 다양한 사람들의 의견이 자유롭게 제기되고 그에 관해 활발하게 토론하며 정책화가 이루어질 수 있어야 하며, 이러한 과정을 통해 궁극적으로 사회적 가치가 창출되어야 한다는 것이다. 오픈 데이터 연구소가 ‘변화 이론(Theory of Change)’에서 개방성과 신뢰성을 강조한 것도 바로 동일한 맥락이라 할 것이다.

셋째, 청소년 빅데이터의 생산적 운용을 위한 빅데이터 컨트롤타워가 필요하다. 호주의 경우처럼 아동 및 청소년을 위한 국가차원의 데이터 거너번스 및 프레임워크 구축, 전략계획 수립 및 빅데이터 연계 방안, 빅데이터 연계에 의한 정책연구, 실행을 위한 분석 및 연구의 촉진, 나아가 아동 및 청소년 빅데이터 활용사례의 공유에 의한 활성화 등을 리더하는 전담기관이 요구된다. 아동 및 청소년 빅데이터 전담기구가 있을 때, 데이터의 원천별 청소년 및 아동의 데이터 구분과(소셜 데이터, 행정/실적 데이터, 조사데이터) 아동 및 청소년의 영역별 데이터 세트의 구성 및 활용 방안을 효율적으로 마련 할 수 있을 것이다.

나아가 빅데이터 전담기구가 중심이 되어 데이터의 제공(provision)과 관련하여 생산된 데이터를 종합적으로 관리하고 개방적으로 활용할 수 있는 데이터 플랫폼 구축과 효율적 관리·운영 체제를 마련할 수 있다. 미국 연방정부는 data.gov 또는 healthdata.gov라는 포털 플랫폼을 구축하고 공공데이터를 제공하면서 정보의 접근성과 활용도를 촉진하기 위하여 많은 노력을 기울이고 있다. 이는 정부 주도의 빅데이터 기반 공공정책을 추진하는 것과 함께, 각종 빅데이터 장비와 소프트웨어들이 주로 미국을 중심으로 각종 기관과 기업 등을 통해 개발·제공되고 있는 것과도 관련된다(이지혜, 재미경, 조명지, 손현석, 2014). 우리나라에도 공공 데이터는 data.go.kr 포털을 운영하고 있으나, 2020년 8월 현재 이곳의 아동·청소년 데이터는 각종 실태연구나 일반현황 관련 자료들을 중심으로 제공되고 있으며, 청소년백서에서 제시되고 있는 내용을 넘어서는 보다 실제적·구체적 자료들은

찾아보기는 쉽지 않다. 연구 자료와 민간기관에서의 자료가 체계적이며 효율적으로 관리되고 있는 경우도 많지 않으며, 관련 데이터를 활용하고자 할 경우에도 접근이 제한되고 승인을 받는 등의 절차적 과정이 선행되기도 한다. 이런 점에서 미국의 “OPEN 정부데이터법”에서 제시하고 있는 “개방적(open), 공공적(public), 전자적(electronic), 필요한(necessary)” 의미를 구현할 수 있는 데이터 제공을 위한 청소년관련 정보를 집대성하는 종합 공공데이터 포털을 개발하는 등의 정책개발이 있어야 할 것이다.

또한 컨트롤 타워가 있을 때 데이터 활용도 제고(promotion)를 위한 연계·공유적 연동 시스템을 구축하는 일이 가능하다. 우리나라에서는 현재 각종 청소년관련 기관을 중심으로 “공공데이터의 제공 및 이용 활성화에 관한 법률”에 따라 데이터 공개와 활용을 위한 노력을 기울이고 있으나 각 기관별로 개별·분산적이며 기관 간 상호 연계적 운영은 결여하고 있다. 대표적인 기관으로 한국청소년정책연구원의 경우에는 패널 조사 데이터를 비롯하여 횡단연구 자료를 공개하고 청소년통계 아카이브센터를 구축하고 있으며, 한국청소년단체협의회, 청소년활동진흥원, 한국청소년상담복지개발원 등의 경우에도 청소년단체/시설 DB, 빅데이터센터, 빅데이터분석 등의 이름으로 관련 자료들을 제공하고 있음을 확인할 수 있다. 그러나, 이들 기관의 데이터 생산과 제공의 경우도 미국의 사례에서 보는 것과 같은 하나의 기관 간 협력체제(예: the Forum)가 주관하거나 또는 데이터 온톨로지나 합의된 지표체계에 따라 지속적으로 공포하는 시스템(예: youth.gov: Child Trend; KIDS COUNT)에 따라 수행되고 있는 것은 찾아보기 어렵다.

더욱이 아동·청소년과 관련한 보건·의료 분야 등의 데이터의 접근이나 정보공유를 위한 노력도 여전히 제한적이다. 이러한 데이터는 개인정보가 중요하고 정보활용에 한계가 있지만, 그에 필요한 제도적 장치를 마련하고 관련 데이터를 공유할 수 있는 방안을 찾아보는 노력이 있어야 할 것이다. 특히 강조할 점은 미국의 경우에서 볼 수 있는 청소년문제 위험예측 대응체제 구축과 관련한 시스템 공유

등과 같은 사례를 우리의 경우는 찾아보기 어렵다는 것이다. 미국의 PSYCKES의 경우, 그것이 빅데이터로서 갖는 의미는 메디케이드 보험청구 시스템을 비롯하여 개인 의료정보를 검색할 수 있는 다양한 전산망과 연동하면서 자살시도자 및 자살 사고자를 관리할 수 있다는 점이다. 우리나라의 경우도 개인정보보호법에 의해 정보공유의 어려움은 있지만, 현행의 정신건강사례관리 시스템과 국민건강보험공단 등의 정보망과 국가응급의료전산망 등이 정보를 공유할 수 있는 연동시스템을 갖출 수 있다면, 자살시도자 및 자살 사고자를 파악하고 예방 관리할 수 있는 방법을 찾을 수 있을 것으로 본다(손해인, 2018).

넷째, 개인정보의 보호(Protection)와 관련한 법적 장치를 마련해야 한다. 특히, 빅데이터 중에서도 가장 많은 부분을 차지하는 비정형 데이터의 경우 스마트폰이나 소셜 미디어를 기반으로 한 다양한 개인 정보를 포함하고 있다는 점에서 해당 정보의 수집과 활용에 앞서 관련 정보의 제공과 관련한 권한과 책임의 문제, 정보유출과 보안의 문제 등에 대한 법·제도적 보호 장치가 선행될 필요가 있다. 실제로 빅데이터 기술이 발전함에 따라 청소년문제의 예방, 발견 및 개입을 위한 기회가 더욱 증대되고 있지만, 그와 동시에 프라이버시와 관련한 함의들이 무시할 수 없는 중요한 이슈로 대두되고 있다.

따라서 빅데이터 전략을 실행할 때는 시민적 자유(civil liberties)를 위반할 잠재성이 따르는 것이 가장 큰 관심사가 되고 있다는 점에 유의할 필요가 있다. 이는 특히 미국의 경우 등록되지 않은 이민자나 동성애자 등과 같은 취약한 구성원에 속하는 청소년들에게 더욱 중요한 문제가 되고 있음을 확인한 바 있다. 예를 들면, 사회적 정서적 스트레스요인을 가진 취약 청소년들의 경우 학교 당국이 자신들을 빅데이터를 통해 모니터링하고 있다는 것을 알게 되면 학교를 떠나는 현상도 있었다(Watson & Christensen, 2017). 이에 따라, 미국의 경우 개인정보보호와 관련하여 미국 연방정부 교육부는 디지털 시대 학생들의 프라이버시(사생활보호)를 보호하기 위한 필수요건(requiremen)과 최상의 실천(best

practices) 기준을 마련하였다. 학생들에게 투명하게 공개하고(transparent), 부모 동의가 적절하다는 점을 고려하고, 학생정보를 보호하기 위해 제정된 FERPA 같은 법에 대한 자각을 유지하는 노력을 하고 있음을 볼 수 있었다.

다섯째, 데이터 생산(production)과 관련하여 양질의 데이터를 체계적으로 구축함은 물론 지속적 자료축적을 가능하게 하는 법적 장치의 확보가 필요하다. 빅데이터의 성공적 활용을 위해서는 수준 높은 다양한 데이터를 축적하는 것이 중요하며, 각종 관련 기관별, 분야별, 영역별로 각기 다른 항목, 형태, 및 구조로 형성되는 자료들을 보다 실제적이고 유용한 데이터가 되도록 해야 한다. 자료를 구축함에 있어서 주제별, 내용별로 체계적인 지표를 통해 일관성 있게 지속적으로 생산하기 위해서는 청소년기본법 등의 관련법을 통해 구체적인 내용을 명시할 필요가 있다. 실제로 우리나라의 경우, 청소년시설의 운영에서 빅데이터를 구축하고 활용할 필요성이 제기되고, 청소년단체와 시설에서는 프로그램에 참여한 청소년들과 수요자들에 대한 요구와 반응을 데이터베이스화할 것도 강조되고 있다(이경상, 이창호, 김민, 2019). 그러나 현실적으로 청소년시설과 단체 수준에서 관련 프로그램과 참여자 등에 관한 정보를 자료화하기는 쉽지 않으며, 이를 위한 기반 프로그램을 만들어 제공하거나 제공자료의 항목, 내용, 형태와 구조 등에 대한 공통적 체계성을 찾거나 데이터 제출을 의무화하는 규정은 없는 실정이다. 앞에서 살펴 본 미국의 경우, 공공부문의 각종 데이터들이 다양한 주제와 형태로 생산되고 있으며 데이터 공개를 의무화하는 법안이 마련됨으로써, 민간부분의 경우에도 여러 가지 형태로 일관적인 항목과 수집방법을 통해 지속적으로 생산되고 있음을 알 수 있었다. 이는 2009년 5월 출범했던 Data.gov의 경우 최초 47개의 데이터베이스로 출발했지만, 11년이 지난 2020년 8월에는 230,000개가 넘는 데이터셋을 구축한 것으로도 알 수 있으며 이러한 발전에는 “OPEN 정부데이터법(OPEN Government Data Act)”이 크게 기여하였다고 본다. 아동·청소년건강관련 공공 데이터로 볼 수 있는 healthdata.gov에서 취합되고 있는 데이터셋의 경우에도

약물남용(substance abuse)과 우울(depression) 등 50여개의 연관주제어를 활용하여 데이터를 검색할 수 있으며, 데이터 포맷도 html, pdf 등 14개 이상의 형태로 제공되고 있음을 알 수 있다. 공공과 민간부문 모두 데이터의 성격은 매우 구체적이고 실제적인 내용을 담고 있는 상황이다. 이런 점에서 우리나라의 경우, 기존의 각종 현황자료나 기초자료 및 연구자료 등이 보다 구체적인 내용을 담고 지속적으로 확보될 수 있도록 형태와 항목 등의 기준을 마련하여 체계적 생산을 담보할 수 있는 제도화가 요망된다.

마지막으로 데이터의 생산·제공·활용 및 개인정보 보호 등에 대한 이해 증진을 물론, 빅데이터와 관련한 우리나라의 국가미래전략을 고려한 청소년분야의 준비(preparation)가 필요하다. 국내에서도 빅데이터 기술 인력을 양성하기 위한 각종 교육 인프라와 실습 기회 등을 다양하게 제공하고 있지만, 청소년관련 기관 종사자와 청소년단체·시설의 청소년지도자 등에 대한 데이터 생산·가공·제공·활용 및 정보보호 등과 관련한 역량함양을 위한 교육기회는 매우 부족한 현실이다. 미국의 경우, 정부주도의 추진전략으로 마련된 “빅데이터 R&D 전략계획”에서 빅데이터 인력을 7대 전략 중 하나로 제시하고 있으며, 4대 빅데이터 지역혁신 허브 중에서 북동부 허브를 중심으로 교육관련 데이터사이언스를 구축하고 있으며, 북동부 학생데이터단(Student Data Corps)를 만들어 전 지역 학생들에게 가르치고 있다. 우리나라에서도 다양한 형태로 유사한 교육이 대학 등에서 실시되고 있으나, 아동·청소년과 관련한 분야에서 이러한 프로그램이 제공되고 있는 것을 찾기는 쉽지 않다. 따라서, 청소년지도자 연수과정을 통해 빅데이터에 대해 공부하고, 청소년단체와 청소년시설 등의 종사자를 대상으로 하는 직무교육과 보수교육 등을 통해서 데이터 생산·제공·활용 및 정보보호의 중요성에 대한 지식과 구체적인 실무역량을 함양하고 전문직 윤리에 대한 인식을 강화하는 방안을 강구할 필요가 있다.

○ ————— 제5장 전문가 심층면접

- 1. 심층면접 개요
- 2. 심층면접에 대한 분석
- 3. 소결

1. 심층면접 개요⁷⁸⁾

1) 조사 목적

본 연구는 ‘아동·청소년 빅데이터 체계 구축 및 활용 방안’을 도출하기 위해 학계, 현장, 행정계 등 전문가를 대상으로 전문가 심층면접을 실시하여 청소년 빅데이터의 정의 및 중요성, 활용 가능한 빅데이터 유형, 빅데이터를 통해 도출해야 하는 사항, 우선적으로 분석해야 할 인과관계 등에 대해 알아보고, 이를 토대로 청소년 빅데이터 체계 구축의 구체적 방법 및 활용방안에 대한 정책방안을 도출하는 것을 목적으로 한다.

2) 연구 대상

연구대상은 아동·청소년 빅데이터 관련 경험이나 전문성을 가진 학계, 행정계, 현장 전문가 10인을 대상으로 구성하였다. 전문가는 교수 및 국책연구기관 소속의 센터장 등 학계, 과장, 부장 등 행정계 및 현장전문가로 구성하였고 학계가 대부분이다. 전문분야는 교육학 3인, 사회복지학 2인, 경제학 및 경영학 2인, 통계학, 청소년학, 사회학 각 1인 총 10명으로 구성되어 있고, 10년 ~ 35년의 경력을

78) ‘제5장 전문가 심층면접’은 한국청소년정책연구원 성윤숙 선임연구위원이 작성하였다.

가지고 있다. 성별은 남성 8명, 여성 2명이며 최종학력은 모두 박사학위 소지자이다. 연구대상자의 인적사항을 정리하면 <표 V-1>과 같다.

표 V-1 심층면접 참여자의 인적 특성

연번	직책	성별	최종 학력	전문분야 및 경력
1	교수	남	박사	교육학(25년)
2	선임연구위원, 센터장	남	박사	교육학(22년)
3	교수	남	박사	사회복지학(20년)
4	연구위원, 센터장	여	박사	통계학(13년)
5	부장	남	박사	청소년활동(21년 5개월)
6	과장	남	박사	사회학(20년)
7	교수	여	박사	교육학(20년)
8	교수	남	박사	사회복지학(35년 10개월)
9	교수	남	박사	경영정보시스템(16년)
10	교수	남	박사	경제학(10년)

3) 조사도구

‘아동·청소년 빅데이터 체계 구축 및 활용 방안’ 도출을 위한 심층면접 질문지를 구성하기 위해 선행연구 분석과 관련분야 4명의 전문가 자문회의를 거쳐 질문지를 개발하였다.

구체적인 질문 내용 구성 영역은 청소년 빅데이터 개념 및 중요성, 국내외 빅데이터 구축 우수사례, 청소년 빅데이터 분류체계 구축 영역 및 세부지표, 빅데이터 활용경험 및 유형, 빅데이터 활용도를 높이기 위한 방안, 청소년 빅데이터 체계 구축 방안, 일반적 특성으로 구성하였고 각각의 영역에 대한 세부적인 질문 항목은 <표 V-2>와 같다.

표 V-2 심층면접 질문지 구성 내용

번호	조사 영역	세부 항목	
1	청소년 빅데이터 개념 및 중요성	빅데이터 정의 빅데이터의 중요성	
2		청소년 빅데이터 개념	
3	국내외 빅데이터 구축 우수 사례	국내외 빅데이터 체계 구축 상황과 우수사례 국내외 청소년 빅데이터 체계 구축 및 활용 우수사례	
4	청소년 빅데이터 분류체계 구축 영역 및 세부지표	청소년 빅데이터를 분류하는 지표 체계에 대한 중분류 및 세부항목	
5		청소년과 관련된 빅데이터 체계를 만들 때 필수적인 세부지표 및 알고 싶은 것	
6		청소년복지 증진 및 정책개발, 미래예측을 위해 청소년 빅데이터를 통해 우선적으로 분석, 도출되어야 하는 인과관계(독립변인, 종속변인) 및 중요하다고 생각하는 사항	
7		빅데이터 체계 구축 중 우선적으로 구축할 주제 또는 중요 사항	
8	빅데이터 활용경험 및 유형	빅데이터 활용 경험(무엇을, 왜, 어떻게)	
9		현재 활용 가능한 정형, 비정형 빅데이터 유형	
10	빅데이터 활용도를 높이기 위한 방안, 장단점, 한계점	정형, 비정형 빅데이터의 활용도를 높이기 위한 방안	
11		빅데이터의 활용도를 높이기 위한 개선점	
12		이용자 입장에서 편리한 청소년관련 빅데이터 체계 활용방안	
13		청소년 빅데이터 체계 구축 및 활용의 단점 및 장점	
14		한국 상황에서 청소년 빅데이터 체계 구축 및 활용의 한계점	
15		청소년정책에 빅데이터를 어떻게 활용할 수 있는지에 대한 의견	
16	청소년 빅데이터 체계 구축 방안 및 발전 방향	청소년관련 빅데이터 관할 컨트롤타워의 필요성 인지 및 이유, 관할 부처 및 기관의 역할	
17		청소년 대상 빅데이터 체계 구축을 위한 선결 사항	
18		빅데이터 체계 구축을 위한 중장기 계획 과정	
19		청소년 빅데이터 체계 구축 및 활용의 발전 방향	청소년 빅데이터 체계 구축 및 활용의 발전 방향
20			
21	일반적 특성	성별, 최종학력, 소속, 전문분야 및 경력 기간	

4) 연구절차

‘아동·청소년 빅데이터 체계 구축 및 활용 방안’을 도출을 위한 심층면접을 실시하기 전에 연구의 안전성과 타당성을 확보하기 위해 연구착수 전에 미리 계획에 대한 심의를 하는 기관생명윤리위원회(IRB) 승인을 거쳤다(202008-HR-고유-017). 심층면접 대상자 선정은 빅데이터 관련 선행연구 수행 경험, 그 분야의 전문가의 추천 및 소개를 받아 선정하는 눈덩이표집(snowball sampling)을 통해 이루어졌다. 심층면접 조사기간은 2020년 8월 20일 ~ 9월 9일 사이에 이루어졌으며, 면접 장소는 사전에 약속한 심층면접 대상자 연구실이나 직장을 직접 방문하여 실시하였다. 심층면접 조사는 질문지를 미리 숙지할 수 있도록 구조화된 질문지를 면접대상자에게 사전에 이메일로 배포하였고, 면접시 구조화된 질문지 내용을 토대로 심층 면접을 실시하였으며, 심층면접 시간은 개인에 따라 1~2시간 정도 소요되었다.

또한 연구자는 심층면접 시작 전에 연구의 목적과 취지, 개인 정보와 관련된 비밀보장 등에 대한 정보를 제공하고, 심층 면접 내용에 대한 녹취를 승낙받아 자료 기록 및 녹음을 실시하였다. 심층면접 실시 후 녹취록을 들으며 스크립트를 작성하였다. 연구자는 질문지 스크립트를 여러번 읽으며, 전문가들 의견의 공통점과 차이점, 중심 키워드를 중심으로 분류화하여 내용을 구조화하여 분석하였으며, 분석 후에 관련분야 전문가들을 대상으로 삼각검증(triangulation)을 실시하여 연구의 신뢰도와 타당도를 확보하였다.

2. 심층면접에 대한 분석

1) 빅데이터 개념 및 중요성

빅데이터가 무엇을 의미하고 빅데이터가 왜 중요하다고 생각하는지에 대한 질문을 하였다. 빅데이터의 의미와 관련하여 응답자들은 기존의 데이터베이스로 처리하기 힘든 방대한 양의 데이터라는 일반적 정의에 모두 공감하고 있었다. 무엇보다 데이터의 양이 작거나 정형화된 형태의 스몰데이터의 반대 개념으로 등장한 비정형 데이터의 등장과 비중 증가는 그러한 빅데이터를 기존 데이터베이스의 처리방식으로는 더 이상 감당할 수 없을 정도로 그 양이 커졌음을 시사한다고 보고 있다.

또한 중요성과 관련해서는 이제 모든 일상이 데이터화 되고 있다는 점, 그로 인한 데이터의 확장과 연계 그리고 데이터 처리 기술의 발전 등 다양한 측면에서 기존의 한계를 탈피한다는 점이 언급되었다. 즉 연구를 위한 자료가 풍부하고 다양해지고, 기존의 전통적 방식을 대체하는 새로운 방식의 자료수집이 가능해졌다는 점, 이를 통해 사회현상, 특히 청소년이나 청소년 활동과 관련된 사안이 보다 명시적으로 드러남으로 인해 이전보다 훨씬 더 세밀하고 정확하게 파악하고 분석할 수 있다는 점, 그리고 구체적인 실증적 데이터를 기반한 이론이나 논리의 발견 가능성, 나아가 사안과 관련한 예측가능성을 지적하였다. 무엇보다 빅데이터의 생산주체로서 청소년이 강조되었다는 점은 청소년 관련 빅데이터 체계구축의 필요성을 내포하고 있다는 측면에서 시사점이 있다.

질문1 : 빅데이터가 최근 사회적으로 이슈가 되고 있는데 빅데이터가 무엇을 의미한다고 생각하십니까? 또한 빅데이터가 왜 중요하다고 생각하십니까?

■ 다양하고 많은 양의 데이터

1. 빅데이터는 기존에 저희가 일반적으로 하고 있는 연구방법이나 데이터베이스로 처리하기 어려울 정도의 방대한 그런 데이터를 의미한다.

4. 데이터가 어떻게 보면 모집단에 가까운 게 많다. 그렇기 때문에 사실 그 자체로 그걸 그대로 써도 된다.

5. 반대가 스몰데이터 인데, 스몰 데이터는 사이즈가 작은 것도 있고 정형화된 데이터. 이걸 이제 우리가 스몰데이터라고 표현하면, 그전에 통계데이터가 다 그런 것이다. 사이즈가 작다는 거는 데이터의 양도 그렇고 그러니까 조사 대상도 한계가 있으니까. 그다음에 이제 정형화된 그래서 몇 명 어떤 생각 하나 해서 우리가 조금 우리가 다 계량화할 수 있는 그런 데이터 스몰데이터라고 한다면 빅데이터는 반대개념인 것.

7. 기존 데이터베이스 관리도구의 능력을 넘어서는 대량(수십 테라바이트)의 정형 또는 심지어 데이터베이스 형태가 아닌 비정형의 데이터 집합조차 포함한 데이터로부터 가치를 추출하고 결과를 분석하는 기술이다. 즉, 기존의 데이터 베이스로는 처리하기 어려울 정도로 방대한 양의 데이터를 의미한다.

10. 어원적으로 많은 양 다양한 종류의 데이터의 집합을 말하는 거 같다. 우리 청소년으로 국한해보면 청소년들과 관련된 다양한 종류의 데이터들의 집합이 빅데이터다 라고 말할 수 있을 것 같다. 그래서 이제 일반적으로 데이터는 앞에 우리가 여기 연구팀에서 정의한 것처럼 데이터의 양하고 데이터의 다양성 또 데이터의 속성 신뢰도 가치 복잡성 이런 것들이 다 포함되어져서 우리가 정의 되어 질 수 있을 것 같다.

■ 기존의 한계 탈피

1. 기존의 이제 연구를 할 때 어떤 뭐 종속변수에 어떤 변인들이 영향을 주는지를 본다면, 어떤 이론적 배경 근거나 아니면 선행연구에 근거해서 이제 자료를 모아왔었는데 이제 그렇게 하다 보니까 활용할 수 있는 변수의 갯수가 되게 제한적이었는데, 이런 빅데이터 기법을 활용함으로써, 그냥 기존에 모아져 있는 어떤 뭐 패널자료나 이런 것들을 가지고 분석을 할 수 있다는 점에서 조금 그 중요성이 사회과학 분야에서도 대두되었다

2. 기존에 한계 있는 계량적 연구나 그걸로는 예측이 어렵다고 느끼는 것 같다. 그래서 그런 면에서 그 빅데이터가 이제 중요성이 높아지고 있다.

3. 사람들의 어떤 어떤 기억력에 의존해서 이제 해결했던 문제들이 사실은 이런 데이터를 통해서 좀 더 명

시적으로 드러낼 수 있게 된다.

5. 빅데이터 시대에는 그렇게 그동안 우리가 알고 싶었지만, 우리의 어떤 조사방법이라든가 뭐 여러 가지 그 자원의 한계 때문에 알 수 없었던 것을 빅데이터 툴이나 여러 가지 이렇게 방법들을 통해서 그런 것들을 파악할 수 있게 되면서 결국 사회현상을 좀 더 잘, 자연스러운 사회현상을 조금 더 정확하게 파악하는 거. 청소년이라든가 청소년 활동이라든가 청소년과 관련된 어떤 그 뭐 이런 부분들을 갖다가 이전보다는 훨씬 더 세밀하게 파악할 수 있다.

6. 예전에는 모집단이라고 하면 그 데이터가 워낙 크기도 하고 그 속성을 잘 모르기도 하고 해서 샘플링도 하고 그랬다라고 하면 이미 행정데이터도 그렇고 텍스트 데이터도 그렇고 비정형 특히 그 뉴스 뭐 댓글 같은 것들 그런 것들 다 그냥 있는 그대로 긁어서 분석을 하는 게 더 낫다.

7. 빅데이터 같은 경우에는 다변화된 현대사회에 상황들을 정확하게 예측해서 효율적으로 작동하게 하고 또 개별 구성원들마다 맞춤형 정보를 제공하고 관리분석을 가능하게 한다는 측면에서 과거에는 이런 것들이 상당히 불가능했는데 이게 이제 현실화되고 있다.

9. 기존의 전통적 방식의 자료수집에 사람들이 직접 물어보는 방식이 점점 어려워졌다. 그야말로 전통적인 방식의 그 개인설문이나 이런 것들이 점점 어려워지고 있고, 기술변화에 따라서 기존의 오프라인에 있는 것에 거기에 ICT기술이 접목되면서 말하자면 새로운 방식으로 자료를 입수할 가능성 그 생산할 가능성이 훨씬 더 높아졌다.

생산하고 저장하는 능력이 가격이 굉장히 싸지면서 기존의 전통적 방식을 대체하는 새로운 방식의 자료수집이 가능해졌고 그것에 대해서 이제

그 가능성을 그렇게 말하면 기존의 통계 생산방식도 전통적으로 기술변화에 따라서 변해왔기 때문에 그래서 이제 큰 데이터가 기존에 알 수 없었던 걸 알게 되거나 혹은 기존에 직접 물어봐서 하던 걸 새로운 방식으로 입수하게 됐다.

10. 사회과학연구는 기본적으로 뭐였냐면 일반적으로 나름대로 정의를 해서 그런 가정을 가지고 관계를 분석하는 연구를 했는데 실제 그런 조작적 정의에 근거한 어떤 이론탐색은 상당히 한계가 있다. 그래서 구체적인 실증적 데이터를 기반으로 어떤 이론이나 논리를 찾아내는 일이기 때문에 빅데이터는 굉장히 중요하다. 특별히 경험적으로 몇가지 하다보니까 그 조작정의에 의해서 이뤄졌던 그동안의 연구들로 밝혀지지 않은 것들을 알 수 있는 좋은 계기가 되더라. 이 빅데이터가.

■ 예측가능성

2. 빅데이터는 인간의 기호, 또는 역사 뭐 이런 거라고 볼 수 있는데, 이런 것들을 통해서 이제 역사를 해석하거나 과거를 돌아볼 수도 있고 그다음에 반복되는 패턴 같은 걸로 미래를 예측할 수 있기 때문에 더 각광 받는 것 같다.

■ AI의 등장

2. 빼지 말아야 할 것은 AI의 등장이다. 그래서 AI가 등장하면서 우리 생활에 지금도 영향을 주고 있고 앞으로도 굉장히 큰 영향을 줄터인데, AI의 핵심이 되는 건 결국은 데이터여서 데이터를 가지고 학습을 해야 하는 거기 때문에 그런 면에서 더 중요해지고 아마 그 중요성은 더 커질 것이다.

5. 인공지능도 빅데이터 기반 인공지능이 만들어진다. 그래서 인공지능

의 3차 부문, 바로 이 빅데이터가 직접적으로 연관돼서 2010년도 부터 대체적으로 됐다

8. 빅데이터 인공지능 뭐 인터넷애 기 많이 하는데 어쨌든 시대적으로 이거 이 부분을 모르고 앞으로 미래를 준비할 순 없지 않은가 그런 생각을 한다.

■ 사회/경제적 측면

2. 사회 서비스 측면에서는 그런 예측을 통한 문제해결과 함께 서비스 향상, 뭐 정책개발 뭐 이런 쪽에 중요한 그 자원으로서 빅데이터가 의미가 있다.

7. 다변화된 현대 사회를 더욱 정확하게 예측하여 효율적으로 작동케 하고 개인화된 현대 사회 구성원마다 맞춤형 정보를 제공, 관리, 분석 가능케 하며 과거에는 불가능했던 기술을 실현시키기도 한다. 이같이 빅 데이

터는 정치, 사회, 경제, 문화, 과학 기술 등 전 영역에 걸쳐서 사회와 인류에게 가치 있는 정보를 제공할 수 있는 가능성을 제시하며 그 중요성이 부각되고 있다. 특히, 기존에 기업이 자체적으로 수집한 자신들의 빅데이터를 마케팅에 활용하는 것 뿐 아니라, 공공 기관에서도 공공데이터를 공개하는 등 개인이 이러한 빅데이터들을 접할 기회도 늘어나고 있다. 실제로 기업에서 빅데이터 활용 방안으로 데이터 시각화에 대한 이야기를 많이 합니다. 데이터 시각화를 통해서 빅데이터를 관리하고 분석하는 즉, 활용하는 개념에 대해서 어렵게 느끼신 분들이 실제로 데이터 시각화를 통해 고민을 해결한 사례를 다수 볼 수 있었습니다. 빅데이터를 활용한 새로운 비즈니스 모델을 만들어 가는 스타트업들도 등장했다.

■ 데이터 처리 기술의 발전

3. 소프트웨어도 그렇고 디바이스

도 그렇고 이런 하드웨어나 소프트웨어 자체가 예전보다 훨씬 더 싸지고 싸진 만큼 빨리 처리가 가능하게 발전이 됐거든요 그렇다보니까 예전에 슈퍼컴퓨터가 지금의 mp3 하나 만큼도 안될 정도이기 때문에 이 데이터를 처리하는 거 자체가 굉장히 좋은 환경이 됐고 그렇기 때문에 거기에 들어갈 데이터의 입장에서 빅데이터가 중요하다.

■ 실시간적인 데이터

2. 이제 모든 일상이 데이터화 되고 있다.

3. 중요한건 이게 얼마나 우리가 실시간적으로 얻을 수 있는 데이터냐 그래서 아웃 오브 타임이 안되는 데이터냐 라는 게 더 중요한 거 같다.

4. 빅데이터는 사전적으로 설계되지 않았지만 그렇기 때문에 오히려

다른 용도로 다른 분석에 쓰일 수 있는 거고, 설계된 데이터들은 대부분 행정자료도 그렇고 이게 이제 통계로 나오니까 얘는 1~2년 시차가 있다. 어떤 거는 지금 이거 추세가 너무 빠르기 때문에 1~2년 지나도 그건 옛날 추세기 때문에 대신에 빅데이터는 그 실시간으로 이제 얻을 수 있다.

■ 데이터 개념의 확장

5. 빅데이터에서 데이터의 개념은 수만 되는 게 아니라 이미지, 영상, 눈으로 볼 수 있는 거 그 다음에 목소리 말로 하는 거 그 다음에 글로 쓰여져있는 거, 숫자로 확인되는 거, 텍스트 이미지 보이스 그러니까 사운드. 숫자 그니까 데이터의 개념이 달라지는 것이다. 이게 다 데이터가 되면서 그러니까 이 데이터는 우리는 그동안은 숫자로만 다 확인했는데 숫자 아닌 것들이 또 굉장히 많은 영역을 데이터로 취급이 되면서 그거를 그 데이터를 언제부터 만들어지기 시작했

느냐 하면은 바로 우리가 인터넷을 접속하면서 인터넷 쓰게 되면서 일단 시작이 됐고, 본격적으로 빅데이터라고 표현된 거는 스마트폰이 중요하게 역할을 생각했다 2009년도부터. 그니까 2009년도부터 스마트폰이 상용화됐는데 이게 모바일이니까 그러면서 스마트폰을 통해서. 같이 나왔던 게 이제 SNS. 검색, SNS 그다음에 스마트폰으로 쇼핑 등 스마트폰으로 수많은 어플리케이션을 갖고 많은 활동하다 보니까 여기에 이제 SNS에서 텍스트를 갖다가 카톡 또는 인스타그램에다가 뭐 좋은 그림 사진을 올리거나 어떤 영상도 올릴 수 있고 그러면서 굉장히 많은 우리가 말하는 그 데이터 그러니까 숫자가 아닌 수많은 데이터들이 이제 점차적으로 많아지기 시작한 거다.

■ 데이터의 연계

6. 데이터들이 연계되면서 점점 그 데이터에 대한 볼륨이 커지고 해서

<p>이제 더 빅데이터라는 뭐 분석 방법도 그렇고 그런 부분에 사람들이 관심을 갖는다. 빅데이터라는 개념 자체가 원래는 현재의 컴퓨터 실력 그 수준으로 도저히 해결할 수 없는 그런 대용량의 데이터를 이야기를 하는데 실제로 그렇지 않잖아요. 어떠한 데이터든 다 요즘에 분석도 하고 뭐 다양한 방법들을 써서 하고 있는 상황이라서 정말 빅데이터가 정말 빅이냐라고 하면 요즘에는 그렇지 않은 것 같다.</p> <p>■ 빅데이터의 생산주체로서의 청소년</p> <p>8. 청소년들이 사실은 이 데이터</p>	<p>생산의 주체이다. 가장 이 기기에 접근 돼 있고 데이터 자체를 많이 활용함으로서 데이터가 생산되는 이런 위치에 있다. 그런 또 사례이기도 하고 그래서 그들이 생각하고 있는 거 아무도 그 활동하고 있는 거 갖고 있는 고민이 모든 것을 파악하려 하면 결국 그 데이터, 빅 비정형데이터에 대한 생각을 좀 많이 할 필요가 있고 정형 데이터 부분도 중요한 것이 과연 우리가 하고 있는 청소년에 관한 일들이 제대로 하고 있는가를 알기 위해서는 현재 하고 있는 일을 잘 정리하고 체계적인 자료들이 좀 많이 확보될 필요가 있다.</p>
---	--

2) 청소년 빅데이터의 개념

청소년 빅데이터의 개념에 대한 수정 및 추가 의견과 관련하여 응답자의 과반 이상은 제시된 개념이 포괄적이므로 조금 더 함축적일 필요성을 지적하였다. 본래의 정의는 기존의 연구를 전체적으로 고려하여 다양성(Variety), 속도(Velocity), 휘발성(Volatility), 규모(Volume), 가변성(Variability), 정확성(Veracity), 그리고 가치(Value) 및 시각화(Visualization)의 8V가 나름 의미 있다고 보고 8v를 모두 포괄하는 개념정의를 시도하였었다.

하지만 응답자들의 견해 중에는 핵심이 되는 3V정도로 정리하자는 의견이 있었고, 먼저 개념정리를 하고 8v는 따로 분리하여 보충설명으로 하면 좋겠다는 의견이 있었다. 나아가 청소년 빅데이터의 개념이 청소년의 부정적 문제에만 국한된 인상을 주므로 좀 더 포괄적인 방향으로 고려해볼 것을 건의하는 의견이 있었다. 나아가 구체적인 수정제안으로 청소년 빅데이터의 본질을 ‘기술 및 능력’으로 보기보다는 ‘정보나 자원’ 정도로 규정하는 것이 더 합리적일 것이라는 의견이 있었다. 본 연구는 이상 응답자들의 견해가 합리적이며 타당한 부분이 있는 것으로 판단되었다. 따라서 전체 의견을 종합하여 청소년 빅데이터의 정의를 다음과 같이 수정하였다.

청소년 빅데이터는 ‘청소년 삶의 제반 영역에서 다양한 형태(Variety)로 빠르게 (Velocity) 생성되는, 대규모(Volume) 데이터 가운데, 청소년이 현실적으로 직면하는 다양한 사안의 적시성 있고 적절한 방향제시를 위해, 유의미한 데이터로 선별, 취합 및 선택적으로 분석되는 모든 데이터’이다.

질문 2. 본 연구에서 청소년 빅데이터를 ‘청소년의 다양한 삶의 영역에서 다양한 형태(Variety)로 빠르게(Velocity) 생성되는, 하지만 휘발성(Volatility)을 지닌 대규모(Volume)데이터 중 유의미한 데이터의 선별, 취합 및 생산적 분석을 통해, 사회가 직면하고 있는, 또는 직면할 수 있는 다양한 청소년 문제의 적시성 있고 적절한 미래지향적 해소가 가능하도록, 가변성(Variability)을 고려하여 정확성(Veracity)을 지닌 가치(Value)로서의 생산적 정보를 도출 및 시각화(Visualization)할 수 있는 기술 및 능력’으로 정의하였습니다. 청소년 빅데이터의 개념에 대한 수정 및 추가 의견 있으시면 말씀해주세요?

■ 부정적인 것으로만 생각할 수 있
으니 더 포괄적인 용어로

1. 이제 이게 청소년 문제라는 것의 그 키워드가 문제라는 게 과연 무엇일까 그게 조금 이제 제가 생각했을 때는 어떻게 보면 예를 들어서 뭐 학교폭력이라던지 아니면 청소년비행이라던가 이런 것처럼 약간 부정적인 것으로만 생각할 수도 있을 것 같다는 느낌이 들어서... 청소년 문제라는.. 그 문제라는 거 자체가 어떤 것을 의미하는 건지 궁금해졌습니다. 그래서 이게 꼭 부정적으로 받아들일 필요는 없는데... 뭐 활동이라던지 아니면은 뭐 여러 가지 문제를 포함해서 저희가 조금 관심 있는 것들을 포괄할 수 있는 단어가 조금 더 적절하지 않을까 왜냐면 뒤에 관심 있는 주제라던가 뭐 이런 걸 하물며 뭐 중분류 소분류에 대해서 뒤에서 읽어보면은 다양한 역량이라던지 뭐 만족도 뭐 이런 것들도 나와 있어 가지고 얘를 조금 더 어 포괄적인 용어면 어떨

까 하는 생각이 좀 들었습니다.

2. 청소년이 수식하는 느낌이 아마 문제라고 하면 정말로 갖고 있는 그 과제나 챌린지가 아니라 이걸 정말로 부정적인 문제가 되는 것을 해결하는 쪽으로 느껴지지 않을까 하는 생각도 드네요.

7. 특히 청소년 연구에서 유용한 가치를 창출하는 데 초점을 두었으면 좋겠다 하는 것

10. 어쨌든 청소년들의 삶을 들여다 보기 위한 청소년관련 모든 데이터를 의미한다고 보면 어떨까

■ 조금 함축적인 용어로

3. 일단 8v라는 말은 일단 좀 빼고 그냥 청소년의 어떤 삶에서 나타나는 이슈들을 사전에 미래지향적 해석을 가능하도록 할 수 있는 기술 및 능력으로 해놓고 여기서 그 하위범주에서

그냥 이런 분류로 이렇게 구분한다가 더 맞는 거 아닌가 싶기도 하고 이게 조금 모든 내용이 다 들어가는 게 이게 정의니까 되게 복잡 했어요 처음에도 어렵다 이랬는데 알고 보니까 이게 8v더라고요

4. 그 빅데이터의 정의가 조금 거니까 여긴 이제 많은 걸..

키워드를 생각하면 대규모 그 다음에 그 가변성 뭐 그런 몇 가지가 특히 중요한 것 같아요

5. 그러니까 따라서 조금 쉬우면서도 좀 이렇게 너무 기니까 이렇게 일반인들이 봤을 때..이 막 틀..맞는 말인데 맞는 말을 좀 쉽게 표현하면 어떻게

8. 그래서 저는 그것이 다 응용버전이고 세가지가 주가 아닌가. 속도 문제라던가 크기 문제라던가 그 세가지 Velocity, Volume 뭐 말하는 세 개 있지 않습니까

■ 표현이 수정 보완되어야

2. 기술이나 능력은 어떤 그야말로 그 할 수 있다 없다는 영역인 거니까 그렇게 해서 그 정보를 도출해가고 시각화해서 미래를 예측하고 문제해결에 활용할 수 있는 뭐 자원 이런..

6. 유의미한 데이터의 선별 취합 및 선택적인 분석 등이라는 표현이 들어가는 게 더 좋을 것 같고 ...생산적인 분석, 생산적인 정보 이게 약간 이 키워드를 뭘로 바꾸는 게 좋을지는 잘 모르겠지만 그냥 정보라고 하셔도 되지 않을까 싶어요 ...기술 및 능력이 빅데이터는 아니잖아요

어쨌든 분석까지 빅데이터로 보고 있다라는 거를 넣고 싶으신거잖아요 그래서 그렇게 마무리를 하시는 게 좋지 않을까 싶은데요.

근데 여기에 들어갈 필요가 있을지를 잘 모르겠어요 안넣으셔도 될 것 같아요 저는 그냥 요즘에 나온 키워

<p>드들만 넣어도 충분하다고 생각을 하거든요</p> <p>9. 근데 제가 보기에는 기본적으로 생산적 가치를 갖는 생산적 정보까지로 정의하는 게 좀 더 상식에 부합되지 않나 그런 생각은 들더라고요.</p> <p>네 그거는 빅데이터를 활용하는 거 관련되기 때문에 저는 그래서 기술 능력 보다는 그냥 그 여기서 생산적 정보까지로 저기 정의하는 게 좀 더 자연스럽지 않나 생각했습니다</p> <p>10. 이거는 그냥 빅데이터에 대한</p>	<p>일반적인 설명에 관한 것이지 우리가 말하려는 청소년을 말하는 빅데이터의 의미는 아니지 않을까 그래서 빅데이터가 특정 그렇게 뭐죠 집단에 한정되기 보다는 전반의 사회적 인식이나 현상에 대한 시사점을 주는 데이터여야 되고 아울러 특히 우리 그 쪽으로 정의를 해보면 제가 정의는 못했는데 어쨌든 청소년들의 삶을 들여다 보기 위한 청소년 관련 모든 데이터를 의미한다고 보면 어떨까 근데 그런 점에서 설명이 돼야 되는데 뭐 양이 많다 이것은 좀 본질은 아니지 않나 그런 의미입니다</p>
--	--

3) 국내외 빅데이터 체계 구축 상황과 우수사례

국내·외의 빅데이터 체계 구축 상황과 우수사례에 대해 응답자들은 국내의 경우 주로 공공 데이터를 언급하였다. 기본적으로 교육부의 에듀 데이터 서비스, 서울시의 버스 노선 및 네비게이션, 그리고 언론 기사 등을 검색할 수 있는 한국 언론진흥재단의 빅 카인즈, 한국정보화진흥원의 빅 데이터 센터 K-ICT, 행정안전부와 정보화진흥원의 공공데이터 포털, 그리고 행정안전부의 국민의 삶을 바꾸는 빅데이터 등을 언급하였다.

국외의 경우는 캐나다의 자살위험 예측 프로젝트 등이 제시되었다. 그리고 빅데이터 국내 우수사례로는 복지 사각지대 발굴 시스템, 위기 아동발굴 시스템 그리고

보건의료 빅데이터가 거론되었으며, 국외 우수사례로는 페이스북과 유튜브를 분석해주는 미국 사이트 빅풋나인이 지목되었다.

다음 청소년 빅데이터 체계 구축 및 활용의 구체적인 우수 사례와 관련하여는 제한적 역량이나 관심으로 인해 소수의 경우만 사례로 제시되었는데 국내의 경우 2013년 여성가족부의 위기 청소년 빅데이터, 교육부의 학생생활기록부 사이트 등이 언급되었고 국외의 경우는 유스 프로그램 데이터 베이스인 best practice가 거론되었다.

<p>질문 3. 3-1) 국내외의 빅데이터 체계 구축 상황과 우수사례에 대해 아는 범위 내에서 말씀해 주세요.</p> <p>3-2) 청소년 빅데이터 체계 구축 및 활용의 구체적인 우수 사례를 말씀해주세요요?</p>	
<p>■ 국내외의 빅데이터 체계 구축 상황과 우수사례</p> <p>1. 한국청소년활동진흥원에서 가장 이제 조금 빅데이터를 구축하기 위한 노력들을 많이 하셨던 것 같고, 그 외에는 이제 교육부에선 그 KERIS를 통해서 지금 에듀 데이터 서비스라고 해가지고 공공데이터를 이제 모아서 인제 연구자들한테 제공을 하고 있습니다.</p> <p>2. 서울시나 이런 데에서 버스 노</p>	<p>선이라던가 그 다음에 그 네비게이션 뭐 이런 것들은 굉장한 빅데이터 체계로 보여져요</p> <p>3. 캐나다 애들 이제 자살위험에 대해서 예측하는 걸 프로젝트로 했어요...웹사이트에 들어가서 이 웹사이트에서 데이터로 분석을 해서 어떻게 분석.. 부정적인 내용이었는지를 확인을 하고 그 IP를 매칭하면 네네이 IP가 어디 주소에 있었는지를 알게 돼요. 그러면 어느 동네 애들이 최소한 부정적인 의의가 많더라던지 아</p>

니면 이 IP로 접속한 또는 이 로그인 데이터로 접속한 사람은 상당히 뭔가 부정적이다 라는 것 같다면 선제적으로 연락해서 케어를 해준다던지 뭐 이런 것도 가능할 것 같은데.

4. 그 이번 주에 그거를 했어요. 차 없는 거리 그러면 지하상가 매출 실적이 카드사 빅데이터로 잡혀있잖아요. 그럼 다음 주에는 하지 않았어요 그래서 그 두 시점을 그거 자료로 비교하는 거죠 실제 그렇게 해서 그 차 없는 거리에 효과 있다는 걸로 아마 나온 거 같아요.

5. 한국언론진흥재단에서 빅카인즈라는 게 있습니다. 빅카인즈. 그래서 주로 언론에 나왔던 것들을 갖다가 검색할 수 있게 그렇게 해놓은 게 있고요 어 과학기술정보통신부하고 한국정보화진흥원에서 하고있는 빅데이터 센터 K-ICT라고 하는데요 그게 있고 행정안전부하고 한국정보화진흥원에서 정보화진흥원이 양쪽

에 다 있는데 공공데이터 포털. 그다음에 행정안전부에선 국민의 삶을 바꾸는 빅데이터라는 게 또 행정안전부 따로 또 하고 있어요 이거는 이제 공공 쪽이고 그다음에 민간 쪽으로 다음 네이버 그다음에 SK텔레콤이 엄청나게 많은 데이터를 가지고 있습니다....이거는 우수사례라고 제가 얘기했는데 나중에 제가 그래서 한 번 빅팟나인이라고 혹시 아시나요? 빅팟나인이 그 미국 사이트인데 페이스북하고 유튜브를 갖다가 분석할 수 있게 해줘요..어 그니까 빅데이터들을 모아놓고 검색해서 자기가 필요한 대로 쓸 수 있는 사이트들이 많아질 거예요...제가 보기에는 우수사례라기보다도 우리는 지금 아직까지 이렇게 많이 쓰이지 않기 때문에 이런 모아놓은 데를 일단 파악하고 거기서부터 어떻게 우리가 데이터를 확보해 나갈 수 있을까를 고민하는 게 중요하지 않을까 생각이 듭니다.

6. 복지 사각지대 발굴 시스템도

이제는 빅데이터 구축 사례 중 우수 사례 중 하나이고요. 그다음에 위기 아동발굴 시스템도 마찬가지예요.... 보건복지 분야에 그 보건의료 빅데이터 정책 심의 위원회가 2000..언제지... 2017년도에 만들어졌는데 보건의료 빅데이터라고 하는 건 뭐냐면 건강보험공단 데이터랑 심평원 데이터랑 암센터 데이터랑 질병관리본부 데이터.. 그 기관 데이터를 연계해서 분석할 수 있는 그 틀을 만들어나가 고자 한 게 보건의료 빅데이터예요... 그 최근에 했었던 정말 빅데이터는 그 복지급여를 받기 위해서는 잘 아실 것 같아요 어쨌든 신청서 내면 그 사람에게 대한 관련된 그런 소득 재산 부채에 대한 정보들 다 끌어오잖아요 끌어왔을 때 그런 데이터들을 다 분석을 해봤었어요 ...캐나다데이터 팝 데이터라는 것도 있고요 그 다음에 제가 지금 캐나다 폼을 쓰고 있는데 그 온타리오주에서 자기네 주민들의 그 어쨌든 저희 건강보험공단 데이터 처럼 그런 데이터들을 쓸수 있게끔

그 데이터 뿐만아니라 다른 데이터 연계해서도 쓸수있게끔 시스템을 만들어 놓은 게 있어요

8. 음 그 외국에서쓰는 말 중에 best practice라는 말이 있어요.... 최고의 실천 최상의 실천을 말합니다 그니까 최상의 실천으로 선정된 프로그램만 모아놓은 유스 프로그램 데이터 베이스가 있어요 그러니까 그런 거 같으면 우리가 좀 이렇게 배울 수 있지 않은가 그런 생각을 합니다

10. 그래도 제일 핫하게 했던 게 서울시 올빼미 서비스 같은 게 그 야간시간에 노선 같은 거 잘.. 이 빅데이터 기반으로 하잖아요

■ 청소년 빅데이터 체계 구축 및 활용의 구체적인 우수 사례

2. 청소년 쪽에 빅데이터 구축과 관련된 외국 사례는 실제로 찾기는 어렵습니다. 왜냐하면 청소년계에서

<p>어 그렇게 빅데이터 체계를 구축할만한 역량이 있는 곳은 아직까지 없었고요 ...청소년 쪽에서는 아직 그 그렇게 우선순위가 지금까지 돌아오지 않았다고 보여지거나 또는 아직까지 관심영역이 어.. 그렇게 그 주목받지 못했던 걸수도 있겠지만 어 국내에서는 그렇게 그 어 청소년과 관련된 빅데이터가 구축되어져있는 사례는 찾</p>	<p>기는 좀 어려웠고요.</p> <p>7. 네 여성가족부에서 2013년에 위기 청소년 빅데이터 분석한 적이 있고요.</p> <p>9. 청소년과 관련해서 가장 큰 건 교육부가 가지고 있는 그.. 학생생활 기록부 그 사이트 있죠 그게 제일 큰 사이트예요.</p>
--	---

4) 빅데이터 활용 경험

빅데이터의 활용 경험과 관련해서는 대체적으로 개인적인 연구 활동이나 정부부처 등의 의뢰를 통한 경험이 다수였다. 연구로는 머신러닝기법들을 적용하여 다양한 패널의 기존 데이터를 진료와 의사결정나무 등의 주제에 적용한 경우, 썸 트렌드 등을 통한 키워드 검색 및 감정분석, 그리고 통계청의 사망신고서 데이터, 건강보험공단 자격정보, 의료이용 히스토리와 직업 관련 히스토리 등 데이터를 merge하여 자살문제를 연구한 경우가 있었다. 정부부처 등 의뢰사업으로는 빅데이터 플랫폼인 센터 구축사업과 그 안에서 혁신서비스로 행정안전부의 안전지도 서비스를 청소년 수련시설이나 청소년 이용시설과 겨울방학 프로그램과 연계한 사업, 그리고 교육복지 우선지원사업 관련 교사 실무자 대상 효과 인식 수준 분석을 통한 교육복지 우선지원사업 방향성과 합의 모색에 참여한 경우가 있었다. 그 외, 다양한 사회문제 해결을 위한 사업에 참여한 것을 확인할 수 있었는데, 응급의료서비스, 보안, 정보시스템, 제품 마케팅, 텍스트마이닝과 여론 등의 주제가 있었다.

질문 4. 4-1) 빅데이터를 활용해 본 경험이 있으신 경우 경험에 대해 말씀해주세요. 4-(1) 무엇을 활용하셨는지?, 4-(2) 왜 활용하셨는지?, 4-(3) 어떻게 활용하셨는지?

4-2) 만약 아직까지 활용해보신 경험이 없다면 어떠한 데이터를 활용해서 어떠한 사항을 분석해보고 싶으신지 의견을 자유롭게 말씀해주세요

■ 연구활동

1. 2016년도부터 조금 이제 관심이 있어서 했었고 그래서 했었던 논문이 한 다섯 편 정도 지금 이제 나왔습니다. 근데 한편은 비정형데이터 이제 테스트망 있는 거를 이용해 가지고 냈었던 건데 이거는 KCI 등재되어있는 논문들 초록을 활용해서 주요 키워드를 활용했었던 연구고, 진로와 관련된 것을 했었습니다. 그리고 네 대학생들 진로와 관련된 것들을 했고요. 그 외의 논문은 뭐 의사 결정나무 뭐 이렇게 해서 여러 가지 정형화된 데이터에 저희가 적용할 수 있는 그런 머신러닝기법들을 이제 적용한 것이 한 네편 정도 있고요. 근데 이거는 이제 저희가 기존에 있었던 데이터들을 활용을 했어요. 예를 들어서 뭐 보시면은 진흥원에 있

는 인적자본 패널이라던지, 또는 한국교육정보원에 청년패널 그 다음에 어.. 진흥원에 고용패널 킵? 그 다음에 작년에 학술대회에서도 발표를 했었는데요 KCIPS 데이터를 가지고 저희가 이제 그 때는 이제 초중학생들 GREET에 영향 주는 변수 탐색 이렇게 해가지고 이제 연구를 했었습니다.

5. 썸 트렌드에 한 번 들어가 보시면 이거 진짜 재밌어요 여기는 주요 키워드 검색뿐만 아니라 감정분석까지 해다 줘요 감성해서. 이게 주로 SNS를 키워드로 하는 거거든요 그니까 국내 다음이 이제 카톡 말고 아마 다른 SNS도 같이 제휴해서 이걸 하는 것 같아요. 그래서 저는 이걸 갖고 학술지..제가 한국 학생에 대해서 지난번에 한국교육학회에서 발표

했었거든요

8. 연구에 관계돼서 필요한 그 키워드 검색을 하게 되고 찾다보면 그런 것이 올라가 있는 데이터 이런 그 저 플랫폼을 확인할 수 있다 뭐 그리고 그런 자료에 들어가서 나온 그 데이터들을 연구에 기본자료로 활용하고 글도 쓰고 할 때 활용할 수 있지 않습니까. 물론 인용도 하지요. 그러니까 이렇게 활용할 수 있다 이런 이런 경험.. 기본이 연구자니까 맨날 이런 거 합니다

9.이거는 통계청이 사망신고서에 있는 데이터가 있고 그다음에 건보공단 자료 자격정보 의료이용 히스토리 그 다음에 그 사람의 직업관련된 히스토리 정보를 쭉...그런 자료를 가지고 분석을 했죠 ... 정형데이터인거죠 그나마 행 정부가 국민들의 건강을 관리하기 위한 의료보험 서비스 차원에서 만들어진 데이터고 사망신고서랑 두 개의 데이터를 병

합해서.. 저희는 아마 그 정도가 실제로 할 수 있는 일이에요. 저도 그 일 한번 해보고 싶었고 제가 찾아갔죠. 복지부로. 이거는 하루에 40명이 아파트에서 뛰어내리는 나라가 이게 정상이 아니다. 이걸 지금도 서른..30 몇 명씩 그렇죠. 하루하루 지금도 누군가가 뛰어내려 코로나보다 더 무서운 거죠.

■ 정부부처 등 의뢰

2. 이름은 키와 빅데이터 센터 라고 하고요. 작년에 과학기술정보통신부에서 국가에 빅데이터체계를 구축하기 위해서 빅데이터 플랫폼 센터 구축사업을 시작하게 됩니다. 총 규모는 10개의 플랫폼에 1개 플랫폼에 10개씩의 센터를 지정을 해서 100개 센터 그 빅데이터 센터를 만드는 거고요. 그 10개의 플랫폼은 그 저기 사전에 주제가 정해져 있어요 ...청소년 쪽을 만들어보려고 타진을 해봤는데 시장자체가 너무 작은 거

조 그래서 가장 그래도 유사한 문화 쪽하고 저희가 결합을 했어요 그래서 아 문화 플랫폼에 한 개의 센터로 들어가는 걸로 그렇게 해서...저희는 보고서를.. 연구보다는요 혁신 서비스를 만들어내야 해요.

작년에 그래서 혁신서비스를 작년에 이 사업이 어.. 거의 8월부터 시작됐어요. 8,9,10,11,12 뭐 이렇게 해가지고 충분하지 않았거든요. 작년에 실험적으로 혁신서비스 만든 거는 세 가지 정도예요.. 청소년 수련시설이나 청소년 이용시설 관련된 문제가 남아있다고 저희가 이제 착안을 하고 실제로 그 행정안전부에서 안전지도라는 서비스를 해요...겨울방학 프로그램을 그쪽 키워드에 맞춰서 그 개발하도록 그렇게 그 하는 거 하나 해드렸고 그 다음에 그 아까 저기 그 안전지도 말씀 드렸는데 그런 안전지도에 청소년 수련시설이 안 들어 있어요. 네 그래서 그 청소년 수련시설하고 이용시설을 집어 넣는 거를 저희가 했었죠.

3. 뭐 여러가지 많이 한거죠. 그래서 뭐 보안도 해보고 정보시스템 쪽도 해보고 뭐 제품 마케팅 쪽도 해보고 뭐 텍스트마이닝을 해서 뭐 사람들 여론 바꾸는 것도 해보고 뭐 누가 뭐 댓글 알바를 하고 있는지를 추적하는 것도 해보고.

4. 교통 분야에서는 로드킬이 큰 이슈예요...그래서 어느 지역이 로드킬이 발생할 확률이 높은가 그거를 추정을 해요 그거를 제가 해준 적이 있어요 ...응급의료서비스에. 근데 이제 저는 119 그거 콜 자료를 가지고 그걸 제가 분석해서 어느 지역에 응급의료 발생 확률이 어느 정도 높으냐...코로나 때문에 예산군 관광에 뭐 미친 영향 해달라 하는데 기존에 정형데이터는 사실은 없죠.

7.비정형 데이터를 분석해본 경험이 있는데 타이틀이 이제 그 교육복지우선지원사업 참여 학생 학부모

<p>사업을 수행하고 있는 교사 실무자들을 대상으로 효과 인식 수준을 분석하고 이를 바탕으로 해가지고 교육복지우선지원사업 방향성과 함의를 좀 모색한 적이 있습니다. 거기에 활용방법으로 이제 그 캡스톤이란 방법을 사용했고 캡스톤마이닝을 분석을 해서 텍스트 네트워크 분석까지 그런 식으로 했습니다.</p> <p>10 저희가 처음으로 국가교육통계정보를 데이터베이스화하는 작업을 해요. 제가 처음 시작을 했어요. 그때는 정말로 불안불안하면서 막 그 데이터를 모으는 작업을 했고 그</p>	<p>게 지금은 발전해서 매년 연례적으로 국가수준에서 유치원부터 초중등 모든 데이터를 모으고 있고요. 그다음에 취업통계 통계 DB 그 다음에 평생교육 통계 DB까지 확장 발전해서 지금 평생교육원 저쪽에서 가져갔지만 그때 이제 그런 일들을 처음 해서 데이터를 모으는 일을 DB를 만드는 일.. 저는 전공이 교육학인데 그런 일을 했어요..</p> <p>■ 기타</p> <p>6. 전 분석방법론 데이터마이닝이랑 방법론들을 다 알고 있어서요</p>
--	--

5) 활용 가능한 정형, 비정형 빅데이터 유형

현재 활용 가능한 정형 빅데이터로 응답자들이 거론한 것으로는 크게 연구기관의 연구 자료들과 정부부처의 각종 데이터들이었다. 특히 정보화진흥원의 공공빅데이터들, 통계청 데이터들, 여가부 청소년 종합 실태 자료, 그리고 청년 패널 등 패널 데이터가 언급되었다. 다음 비정형 빅데이터로는 SNS 소셜 데이터들 중심으로 유튜브를 비롯한 동영상 데이터, 사진데이터, 문자데이터 등 텍스트데이터가 언급되었으며, 상담 데이터라던가 컨설팅데이터 및 카드사 정보 등도 거론되었다. 특별히 청소년 데이터로는 여가부의 청소년 종합 실태와 다른 데이터에

청소년 정보가 들어간 것들로 인구 총조사, 사회조사, 그리고 한국 철도공사의 청소년 할인 정보 등이 언급되었다. 그 외 비정형인 가공되지 않은 원자료를 정형화시키는 경우로 도로공사에 톨게이트 이용정산자료가 거론되었다.

<p>질문 5. 현재 활용가능한 정형, 비정형 빅데이터로 알고 계신 것을 자유롭게 말씀해주세요.</p>	
<p>■ 정형 빅데이터</p> <p>1. 청년 패널이라던지 외에 그니까 개발원에서도 켈스 중단연구를 이제 2005년도부터 하고 있어서 이제 그 자료도 활용할 수 있을 것 같고 또 KCPIIS 데이터, 그 다음에 이제 최근에 각 교육청에서도 자료를 이제 수집해서 학술대회나 뭐 이런 것들 연구자 공모를 통해서 많이들 하고 있어서 이제 그런 자료들을 정형 데이터로 활용을 하면 좋을 것 같고요 ...</p> <p>2. 정형데이터 같은 경우에는 우리가 그 흔히 알고있는 여러가지 통계 데이터 라던가 또는 그 정형화 될 수 있는 여러 데이터들이 있기 때문에 그 이 부분은 특히 청소년 쪽에서는 저희가 갖고있는 e청소년에 있는 활</p>	<p>동실적이라든가 또 시기별 상황이라든가 뭐 이런 것들이 될 거고요. 그 다음에 그 공공데이터포털이나 그 다른 데에 있는 데이터들도 굉장히 주목해서 보고 있는 편입니다.</p> <p>5. 일단 정보화진흥원에서 갖고 있는 거는 이제 얻을 수 있다. 공공빅데이터들을 ...나머지 정형데이터들은 다 각 기관이나 뭐 통계하는 데가 일단 다 갖고있는 거죠. 통계청을 통해서 얻을 수 있는 거겠죠.</p> <p>8. 연구기관에 들어가면 연구자료들 잘 정리돼있고 요즘은 뭐 공개하고 이런 거 다 좋아요. 그리고 뭐 패널 데이터 같은 것도 그렇고 그 다음에 어 뭐 다문화 청소년에 관한 데이터 정책연구자들 그리고 상담 데이터</p>

<p>가 있어요 ...</p> <p>9. 정부부처가 굉장히 많은 걸 가지고 있죠 ...</p> <p>■ 비정형 빅데이터</p> <p>1. 비정형으로는 이제 KCIPS 등 재지 이용해서 초록 이런 거 가지고 했었고요</p> <p>2. 비정형데이터 같은 경우에는 대표적인 게 소셜 데이터인데 ... 상담 데이터라던가 컨설팅데이터 뭐 이런 것들이 그 중요한 비정형데이터라고 보여지고.</p> <p>3. 동영상 데이터, 사진데이터, 문자데이터, 텍스트데이터, 유튜브, 기타 웹소스는 다 쓸 수 있고요.</p> <p>4. 교통 쪽에 5D 자료가 많아서 거기도 일단은 비정형 자료가 많고요. 그러니까</p>	<p>5. 비정형 빅데이터는 활용 가능한 게 사실 다 활용 가능합니다. 안줘서 그렇지 데이터를. 돈 받고 막 다 풀어놔서 그렇지 아까 제가 말씀드렸던 SNS, 쇼핑, 카드사 뭐 등등 해가지고 이런 것들이 다 활용 가능한데 ...</p> <p>7. 이제 SNS 같이 그 비정형 데이터로 이제 데이터를 사용해서 분석하는 것을 의미하는데 이거는 실시간 생성되는 정보를 통해 가지고 더 많은 정보 수집 분석을 할 수 있다는 것이 연구의 특징입니다</p> <p>8. 비정형 데이터는 뭐 저는 유튜브를 주로 뭐 이렇게 활용하고 유튜브 검색 하면 뭐 안나오는 거 없습니다. 웬만한 키워드 검색하면 뭐 다 나오고.</p> <p>10. 말할려면 너무 많지 않아요? 뭐 일상이 전부 다 스캐닝 되고 있는 우리 삶이잖아요. 그래서 뭐 너무 많</p>
---	--

은데...

■ 청소년 데이터

4. 여가부 청소년 종합 실태 그 몇몇 개 누가 봐도 청소년 데이터예요. 근데 이제 다른 데이터에 청소년 정보가 들어간 게 있잖아요. 그러면 그것도 청소년 데이터로 볼 거냐 그렇게 따지면 또 한도 끝도 없죠. 왜냐면 보세요. 인구 총조사에 당연히 청소년 들어가 있죠. 뭐 사회조사 이런 데 들어가 있죠. 아 그러면 엄청 많아요 또. ... 이 자료도 만약에 도로공사에서 만약에 한국 철도공사에서 그 승객이 그 표 예매할 때 청소년은 할인 해주잖아요. 그렇게 되면 이것도 청소년 자료에 들어가야죠. 박사님. 청소년 빅데이터에.

6. 위기 청소년 관련해서 찾아봤을 때 복지부 얘기는 하지만 그 실제 각

각 저희가 정의한 영역별로 이렇게 어떤 시스템 정보들이 있는지를 찾았거든요. 근데 여기에 보면 행정자료예요. 행정자료에서 어떤 자료들이 있을까? 예를 들어서 이 소년 소녀 가장이라던지 위탁가정의 데이터들도 이제 유용하게 쓰일 수 있잖아요. 이런 자료들이.

■ 기타

4. 그다음에 교통 쪽에 가면 이거는 정형 비정형이 약간 경계가 모호한데요. 박사님, 그 도로공사에 틀게이트 이용정산자료가 있어요 그래서 어디서 뭐가 어느 화물 뭐 승용차 어디서 어디 이거 뭐 사실은 그건 비정형이라고 봐야죠. 비정형인데. 그니까 가공되지 않은 원자료를 받아다가 우리가 이제 우리가 맞게 그거를 정형화시키는 거죠.

6) 청소년 대상 빅데이터 지표 체계 분류 영역

(1) 청소년 대상 빅데이터 지표 체계 중분류 및 세부항목

청소년과 관련하여 꼭 들어가야 한다고 생각하는 중분류와 세부 항목에 관한 질의에 대하여 응답자들은 크게 생애설계적 관점과 생애수정적 관점의 두 관점에서 접근하는 모습을 보여주었다.

생애설계적 관점, 즉 청소년의 인생을 생산적이고 긍정적으로 구성하는 측면에 중심을 둔 입장은 중분류로 진로, 주관적 복지감, 심리안전, 인구, 가족, 건강, 교육, 여가, 문화, 시간 이용, 공감역량, 관계 및 사회참여 등등을 언급하였다. 그리고 그에 대응하는 세부항목들을 제안했다. 구체적으로 청소년 개인과 관련하여 홀로 보내는 시간, 독서 스마트폰 사용량과 사용시간, 규칙적인 운동 여부, 아침 식사 여부, 수면 시간, 비만도, 자원봉사활동 등이 있었고 가족 구성 상태에 관련한 것들로 한부모 또는 조손가정 여부, 가구소득 등이 언급되었다.

생애수정적 관점, 즉 청소년에게 부정적 영향을 미치는 것들과 관련해서는 스마트폰 중독이나 인터넷 중독, 분노 조절장애, 공격, 스트레스, 학교폭력, 자살, 불안, 비행, 자아존중감, 자아효능감, 자아탄력성 등이 거론되었고 세부항목으로 비대면 활동에 참여하는 시간, 학교 학원 이외의 하루 활동, 디지털 활용시간 등이 언급되었다. 그 외 기타 의견으로 공간과 시간에 따른 데이터접근의 필요성 제시가 있었다.

질문 6. 청소년 대상 빅데이터에 어떠한 것이 있는지 분류하는 지표 체계를 중분류로 영역화하여 세부항목을 산정한다고 할 경우, 청소년과 관련하여 꼭 들어가야 한다고 생각하는 중분류, 세부항목에는 어떠한 것이 있나요?

중분류: 예시) 인구, 가족, 교육, 역량, 건강, 안전, 심리, 활동, 주관적 복지감 등, 세부항목: 예시) 방과후 나홀로 보내는 시간, 독서율, 비만도

■ 생애설계적 관점

1. 중분류에 진로 관련된 키워드가 들어가면 어떨까 이제 생각을 해봤고 그럼 그 밑에 이제 하위요소로는 뭐 진로성숙도라던지 진로결정효능감이 나 요즘 많이 진로와 관련된 내용들이 있어서 그 부분에 조금 관심 있었고 ... 이 중독과 이제 자기조절 부분이 조금 관심이 있었어요

3 이것도 보면은 뭐 주관적 복지감 심리안전 이런 게 있는데 아래에 보면 나홀로 보내는 시간 이런 것들은 꼭 좋은 의의가 될 것 같아요

4. 제가 생각한 거는요. 중분류를 굳이 하자면 기본적으로 인구 가족

그 다음에 건강, 교육, 여가, 문화, 시간 이용, 그 다음에 주관적 웰빙, 이 정도가 중분류로 청소년 대상으로 하기 때문에. 그래서 이제 인구 가족 중분류에서 그 영역에서는 그 가구 구성 그래서 해당 청소년이 일 세대 같이 동거하는 사람들이.. 그 일 세대냐 그니까 부모 같이 사느냐 아니면 조손 아니 할아버지 할머니 같이 사느냐 그런 거하고 그 다음에 한부모 가정이나 다문화 가정이나 이런 것들이 거기에 포함될 수 있을 것 같고요. 그 다음에 이제 동거 가족 그러니까 이제 그래서 본인하고 부모하고 뭐 분거 하고 있느냐 혹은 엄마 아빠가 맞벌이 떨어져 사느냐 이런 것들도 상당히 중요하지 않아요? 애들이 성장기에 그런 것들이 그 세부항목으로 들어가면 좋을 것 같고요.

5. 인구나 뭐 가족 교육 역량 공감 안전 활동 뭐 등등 그런 것들이 있을 것 같아요 그니까 청소년 관련해서 그니까 관심 있는 거 청소년과 관련해서 이거는 아마 청소년 전문가들이 더 많이 잘 아실 거라고 생각을 하는데 음.. 여기 아 여기 진로도 중요한 게 있구나 진로나 청소년 관련해서 이거는 나중에 제가 그런 것도 있을 것 같고 세부항목으로 어.. 아니 그 제가 왜 독서 같은 거를 얼마나 많이 하는지 그니까 물어보고 싶은 건 되게 많을 것 같아요. 각 영역별로 우리가 스마트폰 얼마나 쓰는지 그 다음에 뭐 비만도 또는 방과 후 나 홀로 보내는 시간 뭐 등등...

7. 중분류에 이제 인구, 건강, 학습 교육, 여가, 경제활동, 안전행동, 관계 및 인식 사회참여 등등..이 기본적으로 이제 통계에서는 사용되는 두 개인 것 같고 세부항목으로는 인구에서는 청소년 연구, 환경연구 뭐 다문

화 학생 건강 분야에서는 초·중·고 학생 비만 통계라던지 규칙적인 운동을 하는 청소년 비율이나 아침 식사 하는 청소년 적정수면 시간을 유지하는 청소년 등등 이런 게 있는데 그 밖에 학습, 교육, 여가, 아까 이제 중분류에 말씀드린 그런 내용들로 그 세부항목들이 좀 추가돼야 된다는 형태로 제가 설명하겠습니다.

8. 시민사회 역량 중에 자원봉사활동에 참여시간 자원봉사 활동에 참여했던 기관의 수 그리고 뭐 시간이 문제가 아니라 그 뭘니까 빈도의 수 라든가 이런 다양한 것이 있어요. 근데 이제 우리도 그렇게 뭐 자원봉사는 자원봉사대로 데이터가 만들어지고 있지 않나 생각이 드는데 그런 하위 세부 분류가 그런 게 좀 더 나가서 네 여러가지 다양하게 좀 이렇게 분류할 수 있는 기준이 있으면 좋겠다 그런 생각이 들어요.

■ 생애수정적 관점

1. 이제 중독과 관련된 건 뭐 스마트폰 중독 뭐 예전에 인터넷 중독 이제 뭐 좀 그런 쪽에 좀 관심이 좀 있었고 ...최근에 많이 뭐 분노 조절장애라던지 그런 부분들이 의외로 많다고 생각이 들어서 드러나지 않았을 뿐, 그런 부분들을 혹시 어.. 알 수 있을까 하는 그런 궁금증이 조금 있습니다.

2. 세부항목 같은 경우에는 ... 최근 들어서 코로나 상황을 얘기 안 할 수 없는 거니까 그 비대면 수업이나 비대면 활동에 참여하는 시간 그 다음에 저희는 그 뭐 이 상황에 관심도 그렇습니다만은 청소년들이 하루를 어떻게 사느냐가 제일 궁금해요. 네 그래서 학교, 학원 이외의 활동이라든가 학업 이외의 활동장소라든가 뭐 이런 거 항상 궁금해하고 있고요. 그 다음에 그 디지털과 관련된 활용시간이 얼마나 되는지 그니까 예를 들면 게임시간 이런거 말고 디지털을 활용

해서 뭔가하는 시간 또 디지털디바이스를 활용해서 활용하는 시간이 얼마나 되는지 그리고 어떤 디바이스를 많이 이용하는지 그 다음에 웹이나 앱에 머무르는 시간이 얼마나 되는지 뭐 이런 것들을 좀 알 수 있으면 저희가 뭔가를 서비스 하는데..

6. 가족의 속성을 가지고 있는 정보가 있고 아동 개인.. 그니까 청소년 개인에 대한 속성을 가지고 있는 정보가 있잖아요. 그거를 구분을 하셔야될 것 같아요. 그니까 물론 당연 세부지..세부항목들은 다 지금 나열을 해 주신거는 아동 개인에 대한 속성들이지만 어쨌든 청소년이 뭐 위기청소년이 됐든 뭐가 됐든 그 가족이란 울타리 안에 있잖아요. 그래서 가족의 속성을 나타내는 아까 그 제가 말씀드린 소득이라든지 네네 그런 정보들도 여기에 담겨져 있어야 되질 않을까 싶어요. 그거는 제가 고민을 했던 부분이에요. 이 아동통계 행정패널 즉 그 데이터로 뭔가를 작업을

할 때..경제적인 수준이 어쨌든 가구.. 가구소득 들어갈거고

제가 드리고 싶은 말씀은 가구 환경에 대한 속성도 필요하다라는 말씀을 드리고 싶은 거예요. 왜냐면 가족 가구 유형도 물론 그 중의 하나지만 어쨌든 그 청소년을 구성하고 있는 주변의 그런 환경 자체도 중요하잖아요. 그리고 아 지역사회 관련된 정보도 저는 필요하더라라고 생각을 했었던요.

10. 공격, 스트레스, 학교폭력, 자살 저는 뭐 학교폭력 그렇죠. 스마트폰 중독, 인터넷 중독, 불안, 비행, 사회불안, 문제행동, 위험요인, 충동, 또 때로는 일부는 비만도 있대요. 요즘 이런 것들.. 아동학대, 사회적 위축, 게임중독 뭐 방임 별의별 많아서 이럴 거

같고 그 다음에 이것으로 뭐 자아존중감, 자아효능감, 자아탄력성...

■ 기타

9. 이게 기본적으로 두 가지에 그 생각이 있는 거예요. 하나는 뭐냐면 공간이라는 개념에서 생태학적인 접근을 하는 거예요. 브론펜브레너의 접근이죠. 그니까 말하자면 개인이 있고 가족이 있고 그 다음에 또래 집단이 있고 지역사회가 있고. 그 다음에 국가 이제 거시적인. 그래서 그 삶 자체의 맥락을 제공하는 것과 애들의 생활 이게 라이프가 중요한데 이걸 이제 그렇게 포착하는 공간적인 데이터 접근이 하나 있고 그다음에 애가 시간에 따라서 변하니까 시간에 따라서 타입이 하나..

(2) 청소년 빅데이터 체계구축 필수 지표

청소년과 관련된 빅데이터 체계를 만들 때 꼭 들어가야 할 세부지표에 관한 질의에서도 응답자들은 크게 생애설계적 관점과 생애수정적 관점의 두 관점에서 접근하는 모습을 보여주었다. 전자는 진로, 직업 등, 그리고 관련하여 청소년들의 일과시간에 대한 관심이 있었고, 특히 긍정적 지표 개발의 중요성을 강조하기도 하였다. 생애수정적 관점의 지표로 생각하는 키워드로는 비행행동, 가출, 자살, 학교 밖 청소년, 미디어중독, 청소년 정서 행동 장애, 가정폭력, 부모자녀갈등, 소비문제, 신체적·정신적 건강, 무엇보다 비교의식 등 청소년 관련 대부분의 문제가 거론되었다.

질문 7. 청소년과 관련된 빅데이터 체계를 만들 때 꼭 들어가야 할 세부지표, 알고 싶은 것은 무엇입니까?	
<p>■ 생애설계적 관점</p> <p>1. 6번이랑 조금 비슷한 것 같아요... 진로 관련해서는 요즘 학생들이 뭐 대학교에서도 졸업하고 나서도 직업이라던가 이런 것들 하고 좀 연결돼 있고, 진로 뭐 관련된 정책들도 2009년도부터 계속 나오고 있었는데 거기에 대한 뭐 실효성이라던가 정책적인 것에 점검을 위해서도 좀 필요하지 않을까.</p>	<p>2. 청소년층과 관련된 그 데이터들을 좀 알고 싶다는 것을 어 계속적인 욕구지요 ...청소년들의 일과시간 같은 건 저희가 아까도 말씀드렸 다시피 지속적인 관심이에요.</p> <p>8 저는 positive한 어떤 그 발달 성장 이런 거에 대한 걸 우리가 먼저 정해놓고 그것이 과연 그렇게 가고 있는가를 보는게 되게 중요해요. 그래서 뭐 보고서에도 뭐 부정적인거 긍정적인 거 얘기하는데 우리가 일반적으로 부정적인 항목들에 강해요. 부정적 항목에 관심이 많아요 그</p>

리고 그쪽은 많이 개발돼있어요. 근데 긍정적 지표가 부족해요 ...그냥 혼재되었는 이런 아까 말씀드린 대분류, 중분류, 소분류 이렇게만 한단 말이에요. 그걸 청소년의 긍정적 발달관계된 부분과 부정적인 요인, 부정적인 상황, 이걸 구분해서 하면 긍정적인 것은 긍정적인 대로 더 촉진시키고 보호요인 하잖아요. 부정적인 건 부정대로 또 우리가 더 안생기도록 이렇게 할 수 있는 어떤 기준이 마련된다 그래서 이 세부지표 만들 때 그런 생각을 넣고 아예 그거를 구조화하면 좋겠다... 이상적으로는...

■ 생애수정적 관점

3. 저도 사실 비슷해요. 그냥 비행 행동, 가출 아니면 애네들이 왜 학업에 집중 못하는가 학교에 대해서 도 대체 어떻게 생각하는가 옛날이랑 좀 많이 틀리잖아요 우리가 막연히 생각할 때 요즘 애들이 학교를 안 좋게 생각, 그냥 막 한다고 생각할 수도

있다 이런 얘기도 꽤 많은데 정말 그러나 궁금한거죠 ...근데 만약에 세부지표를 넣는다면 아마 그런 것들 또 자살. 자살이 저는 제일 시급하다고 생각해요. ...그거 아니면 가정폭력 이잖아요. 그니까 그런 키워드에 맞는 지표가 생기면 좋을 것 같아요 그래서 위험수준 뭐 양호수준...

4. 어 저는 그 주관적 웰빙 영역에서요. 학습활동 관련 부모하고 그 청소년하고 갈등이 어떻게 되느냐. 이런 거가 만약 이 데이터로 오히려 저는 특히 단순히 갈등에서 끝나지 않고요. 박사님 학업성취까지. 그래서 부모가 학생을 그냥 놔두는 게 애가 주관적으로 행복하고 장기적으로 학업성취도 긍정적이더라 라는 게 보여지면 좋을 것 같아요.

5. 소비, 건강상태, 신체적 건강, 정신적 건강상태 요즘 건강이 그렇게들 나빠지고 있어요. 애네들도 약간 좀 정신적으로 힘든 애들이 좀 많

이 늘어나고 있어 가지고 여기도 좀 주시를 해봐야 될 것 같아요 ...그냥 가치 어떤 가치관? 이게 그런 건 쉽지 않는데 관심사가 뭐고 꿈이 요즘 애들 되게 없거든요. 장래희망. 어 이런 거 그 다음에 생활시간. 그니까 청소년 생활시간 조사를 해요. 그것도 의미가 있는데 한 번 만약에 빅데이터를 이용해서 만들면 되게 중요할 것 같아요 생활시간이.

6. 제일 먼저 일단 연구적인 측면에서 연구가 필요한 부분이 위기 청소년 부분이잖아요. 도대체 어떤 이유 때문에 이런 애들이 무슨 생각을 가지고 그리고 이 아이들에게 필요한 지원이 무엇인지 이미 다 연구는 하셨겠지만 애를 데이터로 한 번 쪽보는 것도 저는 필요하다고 생각하거든요. 지금 오픈되어있지 않은 데이터들로 여기 행정자료들 연계해서. 그리고 정신건강 증독도 그렇고 그것도 하나 일 것 같고 그 다음에 이 방과후 나 홀로 보내는 시간에 뭐

를 하는지에 대한 것도 그니까 SNS를 하는지.

7. 이거는 네. 청소년 문제 관련된 거가 대부분 들어가야 된다고 생각을 했는데 위기 청소년 학교 밖 청소년 미디어중독청소년. 정서, 행동, 장애 청소년 어..학교폭력청소년 또 그 밖에 청소년 활동과 관련된 거 이런 것들이 좀 필요하다.

9. 우리나라 애들의 가장 큰 어려움이나 이런 건 뭐냐면 제가 보기에 비교예요, 비교. 비교에서 오는 거예요 근데 엄마가 비교를 하죠. 우리나라 자체가 사실은 굉장히 집단주의적인 비교문화를 가지고 있는 나라라서 근데 그리고 그 ...주관적 행복감 그거를 낮게 하는 데 있어서 애들이 그 음.. 이거에 대해서 아주 비판적인 사람들은 몇 백년이 지나도 안 바뀔거다 이런 사람도 있긴 해요. 굉장히 이게 제가 볼 때 그 우리가 물질적 결핍이 그렇게 심한 나라가 아니

잡아요. 근데 결핍은 결핍이 아니라 비교에서 오는 거예요. 계속 비교해서 그냥 그렇게 살아가면 문제 없는데.. 너무 피곤하고 그래서 이 집단 주의적인 사고방식이나 이런 새로운 그런 가치관 그런 쪽에 좀 더 많은 걸 그런 걸 부각시켜서 그래야 된다고 생각해요 ...

10. 빅데이터가 만들어지면 세 개의 그룹이 쓸 수 있을 것 같더라고요. 우선 정책연구자 또 일반적인 저처럼 공부하는 학자들 그렇죠? 그 분야

에. 그 다음에 그걸 기반으로 무슨 서비스나 뭐 하는 일반 유저들. 제가 깜짝 놀랐던 게 그 KEDI에서 그 통계정보 이거를 사업할 때 놀랐던 거요. 학생들 키나 몸무게를 알고 싶은 사람들이 꽤나 있더라고요. 너무나 놀랐어요. 제가 그런 일 하기 전에 전혀 누가.. 왜그럴까 했더니 교복 만드는 사람들. 그게요. 전혀 아무것도 아닌거 같은데, 책상 만들어야지, 이게 다 걸리더라고요. 그러니까 그게 의미있는 데이터인거예요 실은.

(3) 청소년복지 증진 및 정책개발, 미래예측을 위해 청소년 빅데이터를 통해 우선적으로 분석, 도출되어야 하는 인과관계 및 중요성

청소년복지 증진 및 정책개발, 미래예측을 위해 청소년 빅데이터를 통해 우선적으로 분석, 도출되어야 하는 인과관계(독립변인, 종속변인)에 대하여 면접자들은 독립변인으로 미래나 삶 등에 대한 가치관과 그 외 일반적인 독립변인으로 많은 키워드를 제시해주었다. 지역, 성별, 연령, 학년, 학교, 사교육, 가정 배경, 가족과 관련된 부모 사별, 부모 재혼여부 등이 있었고, 종속변인과 관련하여 부정적 경험과 문제 관련 키워드가 많았는데 구체적으로 나열하면 임신, 낙태 중절, 성추행, 성폭력, 성관계, 우울증, 애정 결핍, 수면제, 정신치료, 자살상담, 항우울제, 경제적 어려움, 실직, 기초생활수급, 이성 관계 어려움, 인터넷중독, 장애인, 질병,

외모, 분노, 스트레스, 불면증, 충동행동, 가출, 폭력, 퇴학, 자퇴, 자해, 음주, 흡연, 약물, 본드, 사용, 위기 행동, 범죄 행동 등이 거론되었다.

<p>질문 8. 청소년복지 증진 및 정책개발, 미래예측을 위해 청소년 빅데이터를 통해 우선적으로 분석, 도출되어야 하는 인과관계(독립변인, 종속변인)에는 무엇이 있을까요? 중요하다고 생각하는 사항을 말씀해 주세요.</p>	
<p>■ 일반적 키워드</p> <p>1. 종속변인은 아까 말씀드렸던 그 두 가지 진로랑 그다음에 중독과 자기조절에 조금 저는 관심을 두었고요.</p> <p>2. 청소년 쪽에서 구할 수 있는 데이터와 그 외 일반 소셜데이터들을 같이 이제 매칭해서 보면 일이 왜 그렇게 벌어지는 지에 인사이트를 받을 수 있을 것 같아요.</p> <p>3. 뭐 미래에 대한 가치관, 삶에 대한 가치관, 친구에 대한 가치관, 돈에 대한 가치관 뭐 이런 게 있는데 그런 가치관들을 가지고 사실은 뭔가 분석을 해야 되는 거죠. 하게 되면. 그게 사실 인과관계로 보여줄 수 있는 개념</p>	<p>이 될 것 같아서. 그래서 그런 독립변인들이 필요하다 생각하고 종속변인은 좀 전에 말씀 드린대로 자살이나 아니면 가정폭력 이런 부분에서 대체 왜 애네들은 신고를 하지 않느냐 그걸 생각해보면 그런 가정폭력 자살 뭐 이런 종속 변인에 대해서 관련된 독립변수들 찾는 게 사실 우선적으로 중요하다 라고 생각이 들어요.</p> <p>4. 가장 중요한 청소년의 주관적 웰빙이죠. 그거에 대한 결정요인 그 이제 주관적 웰빙이라는 게 워낙 다차원적이고 모든 요소들이 다 개인의 주관적 웰빙에 대해서 영향을 미치잖아요. 행복에 대해서요. 그런데 그렇게 될려면 기본적으로 그 청소년 개인에 대한 속성이 통제돼야 되고 그리고 청소년의 가구 부모라든지 이런</p>

것이 통제되어야 되고 그다음에 이제 학교 있잖아요. 학교 속성도 통제가 되어 되고 이런 것들이 다 통제되어야 되기 때문에 이게 이제 연구를 위한 특수목적 계획된 조사가 아니면 사실은 여기저기서 데이터를 갖고 와야 되고요. 그거 다 합해서 해야 되고 그렇게 될려면 여러 종류의 빅데이터만 가능하고 그거를 병합만 할 수 있다면 ... 근데 내가 특히 보고 싶은 건 뭐 해당 청소년의 속성이나 가구 속성 학교 속성을 통제했다는 전제 하에 특히 부모의 속성이 있잖아요. 교육수준 직업, 소득, 뭐 맞벌이 여부 이런 거에 따라서 자녀의 그 학업적 성취 있잖아요. 그거에 대한 기대하고 이로 인한 갈등이 과연 청소년의 주관적 웰빙에 뭘 미치느냐 그런 거를 좀 해보고 싶은 게 제 개인적인..

5. 제가 독립변인으로 생각해봤던 거는 이제 지역이 이제 중요하고 성도 중요하죠. 나이도 중요할 것 같고 학년도 학교면 학년도 좀.. 그 다음에

학교 그 다음에 이제 사교육은 종속이나 독립 다 될 수가 있는데 사교육 굉장히 중요한 영향을 미치니까. 그 다음에 가정 배경 이런 것들이 일반적인 독립변인이긴 해요. 근데 이런 것들이 또 중요하죠 그다음에 종속 변인으로는 뭐 아까 제가 말씀드렸던 소비나 독서, 건강, 성적, 뭐 뭐 그 다음에 장래희망 등등...

8. 18세 선거권이 생겼잖아요 이게 1년 됐습니다. 그런데 음 과연 그 것이 어떤 의미가 있는가 나타나고 있는지 모르겠어요 ... 학교에서 어떤 청소년들의 리더십의 수준이라던가 리더십의 경험이 투표 행태에 어떤 영향을 준다는 거 뭐 이런 거. ...또 그 멘토링이 있어요. 멘토링도 우리가 많은 연구들이 있습니다. 멘토링 효과에 관한 연구들이 많은데 너무 사업들이 많아요. 유사 사업들이 부처마다 기관마다 여러가지 형태로 또 사업의 종류별로 예산 베이스에 따라서 많습니다. 전체를 비교할 수 있는

이런 것도 안필요한가 싶어요.

9. 가치관 관계 이런 쪽 좀 더 좀 관심을 가지고 그리고 행복 이런 거. 그 우리나라가 청소년들이 주관적 행복감이 낮다는 건 잘 알려져 있는데 그러면 그것이 왜 낮고 그거에 대해서 우리가 뭘 해야 되는지 그거는 굉장히 중요한 문제인데 우리가 그냥 분석만 하고 실제로 무언가 다루진 못하는 점.

10. 소년 요즘은 이제 복지 이 관계 저는 좀더 있었으면 좋겠다고 생각하고 과거에는 청소년 교육에 많이 있었잖아요. 뭐 저기 학업이나... 청소년기도 삶에서 중요한 시기니까 그 자체로 의미 있게 미래를 준비하는 일종의 또 조금 힘들어도 괜찮아 미래준비하는 거야 이러지 말고 그 시기를 저는 인정을 했으면 좋겠고 다만 이런 관계 속에서 인과관계에 관한 이야기는 저는 이렇게 이해하고 싶더라고요. 빅데이터가 만들어지면

저는 인과관계에 관한 독립변수는 훨씬 더 자동적으로 많이 생길 것이다.

■ 부정적 키워드

6. 일단 그 위기청소년 왜 위기청소년이 되었는지 ... 예를 들어서 한 부모 한부모 가정이다 라고 하면 한 부모 관련된 그런 프로그램들을 얘기 그런..약간 정신건강 그니까.. 이 어쨌든 올바르게라고 얘기를 하지만 어쨌든 뭐라고 설명을 드릴까요.. 건강하지는 않은 거잖아요.

7. 종속 변인은 아까 말씀드린 그런 문제행동을 갖고 있는 청소년의 주요문제와 관련된 내용들. 예를 들어서 정신건강, 자살, 집단따돌림, 폭력 행동, 우울 행동, 학교부적응, 금지 행동 등등이 들어가야 된다고 생각했고 독립변수로는 이제 여기에 관련된 주요변인들 뭐 가족과 관련된 부모 사별, 엄마 사별, 아빠 사별, 새엄마, 새아빠, 계모, 계부, 엄마 재혼,

<p>아빠 재혼, 임신, 낙태, 중절, 성추행, 성폭력, 어 성폭행, 성관계 뒤 정신장애, 우울증, 애정 결핍, 수면제, 정신 치료, 자살상담, 항우울제, 자살 위험성 또 친구자살, 연예인 자살, 경제적 어려움, 실직, 기초생활수급 또 생활</p>	<p>고, 이성 관계 어려움, 인터넷 중독, 장애인 질병, 외모, 분노, 스트레스, 불면증, 충동행동, 가출, 폭력, 퇴학, 자퇴, 자해, 음주, 흡연, 약물, 본드사용, 위기 행동, 범죄 행동 등등.</p>
--	--

7) 청소년 대상 빅데이터 활용 방안

(1) 정형, 비정형 빅데이터의 활용도를 높이기 위한 방안

청소년 통계활용에서 정형, 비정형 빅데이터의 활용도를 높이기 위한 방안으로 응답자들 중 다수는 전처리의 중요성과 데이터 신뢰의 문제 그리고 데이터 개방을 지적했다. 먼저 비정형 데이터와 정형데이터 둘 다 데이터 전처리를 위해 시간과 비용이 많이 들어 효율성이 떨어지는 문제를 지적했다.

따라서 데이터 클리닝을 비롯 연구자 편의를 고려한 기초 가공이 이루어진다면 빅데이터 활용도가 상승할 것으로 진단했다. 다음으로 데이터 신뢰의 문제를 언급하였는데 특히 비정형데이터의 경우 일반화의 어려움과 타당성 있는 설명력의 결핍 가능성이 언급되었다.

따라서 신뢰성과 타당성 있는 빅데이터에 대한 주기적인 발표의 필요성을 제기하였다. 그리고 마지막으로 데이터 개방과 관련하여 필요한 자료에 대한 홍보 및 접근성이 높아져야 한다고 지적하였다. 데이터의 실질적인 적합성과 실효성을 높이기 위해서는 수용자가 필요로 하는 좋은 품질의 자료를 언제든지 검색할 수 있어야 한다는 것이다. 그 외 활용도를 높이기 위한 방안으로 풍부한 데이터의 양과 사용자의 고양된 데이터 마인드셋이 필요하다는 지적이 있었다.

질문 9. 청소년 통계활용에서 정형, 비정형 빅데이터의 활용도를 높이기 위한 방안은 무엇이라고 생각하십니까?

■ 전처리의 중요성

1. 데이터 전처리에 시간과 비용이 되게 많이 드는 것 같아요. 여기 이거 어 비정형 데이터도 마찬가지로 정형 데이터도 그런데... 키워드 위주로 하다 보니까 한글의 어떤 특수성인데 뭐 띄어쓰기라던지 이런 부분들이 상당히 어 그것을 다 일일이 다 손으로 정리를 해줘야 해요 ... 정형데이터 같은 경우에도 이미 기존에 뭐 기관에서 자료가 모아져 있지만 그 자료들마다 데이터를 처리하는 방법들이 다 달라요... 이런 데이터들을 다 하나씩 다 이렇게 뭐.. 저희가 이렇게 데이터 클리닝을 해줘야되는 그런 부분들이 좀 많이 시간이 걸렸던 것 같아요.

4. 그 다음에 워낙 방대한 자료가기 때문에 연구자 편의를 고려한 기초 가공은 좀 이렇게 해주면 빅데이터는 활용도가 훨씬 높아질 것 같은

데요.

5. 전처리 무지하게 중요하죠 만약에 누군가가 전처리를 다 해준다 그럼 활용도는 상당히 높아질 것 같아요. 많은 사람들이 자기가 필요에 따라서 가공해 쓸 수 있게 될 테니까 ... 전처리를 누군가 해주면 도움이 되는 건 확실하거든요. 그거는. 그게 제일 큰 어려움이니까.

10. 전처리 과정이 중요하다고 봐요. 이 전처리가 잘 안되면 수없이 갖고 있는 자료 손 댈수가 없어요

■ 신뢰의 문제

5. 두 가지로 봤는데요 일단 신뢰로워야 된다. 데이터란 이제 신뢰성을 이제 확보해야 되.. 근데 비정형 데이터를 확.. 하나 비정형데이터의 특징 중의 하나가 이게 일반화시키기

가 어려운 특징이 있습니다. 그니까 정형데이터는 우리가 모집단에서 샘플을 표집 해가지고 전체를 설명할 수 있는 설명력을 갖고 있는데 비정형데이터 같은 경우는 그냥 특정 집단이 생각하고 있는 예를 들면 노인들은 안하잖아요 SNS.... 그거 자체가 뭐 이게 일반화시켜서 그렇게 완전히 하면은 위험.. 안된다는 게 이제 제 생각이거든요 ..두 번째로 주기적으로 발표돼야 된다. 주기적으로 이게 빅데이터 흐름이라고 하는 거 자체가요. 굉장히 의견이 사람마다 굉장히 유행을 타는 어떤 특성들이 있어서 주기적으로 같은 현상에 대해서 어떤 생각들 어떤 것들이 표현되는지 이런 것들을 봐줘야 아.. 이게 어 그니까 지금 제대로 좀 더 많이 쓸 수 있는 그래서 신뢰롭고 신뢰로운 데이터 주기적인 데이터를 만들어내는 게 중요하지 않나 라는 생각이 들었습니다...

9. 저는 제일 큰 문제가 전통적인 문제인 신뢰성과 타당성의 문제예요.

■ 데이터 개방

2. 두번째로는 수요자의 요구에 맞는 데이터를 제공할 수 있어야 합니다.

4. 이용가능한 정보에 대한 홍보? 그 다음에 이제 접근성? 개선 실제 연구자들이 쓸 수 있게 그런 게 필요하고요. 그 다음에 또 하나는 특히 빅데이터이기 때문에 통계자료에 대한 통계 품질이 있잖아요. 그게 쉽지가 않아요...

6. 데이터 개방. 방법론이야 워낙 다른 쪽에서도 하고 있으니까 그냥 그런 연구들을 진행을 하면 될 것 같고요. 저는 일단 어떤 데이터들이 있는지에 대한 파악도 안되고 있으니까 특히 청소년 쪽에서는 그래서 저는 이 연구가 되게 중요하다고 생각을 해요. 파악부터 해야 뭔가 다음 스텝을 넘어갈 수 있는데 지금 그런 상황도 아니니까.

8. 청소년관련 통계에는 뭔가 그 우선 뭐 우리가 뭐 이렇게 보호해야되는 정보가 있을거니까 그 범위 밖에서는 누구든지 좀 이용할 수 있어야할거 아니냐 하는 거지. 우린 여전히 무슨 이유인지 조금 더 통제성이 있어요. 그리고 아무나 접근할 수 없게 돼있는 것도 있고 그래서 미국에 그 정보공개 법이라고 있습니다 ... 개방적이고 대중 아무나 접근할 수 있으면 전자화된 꼭 필요한 데이터 이런 뜻입니다 ... 공공데이터는 아무나 이용할 수 있잖아요. 아니 제 말은 데이터의 이용이 아무나 관심을 가질 데이터들이 만들어져야죠. 자기들이 행정적으로 필요한 데이터만 만들어 놨잖아요 ... 그러니까 적합성 실효성 있는 이렇게 아까 necessary라 했잖아요 그게 누구의 필요한 말이에요. 청소년들이 필요하는 자료가 있다면 청소년이 검색해서 나와야 되죠. 근데 정부의...활용도를 높이려고 그러면 진짜 OPEN 돼야 되죠. 얼마나 전자화 됐는가 얼마나 꼭 필요한 자료인가.

■ 기타

2. 데이터 활용도를 높이기 위해서는 일단은 양적으로 데이터가 제공되는 데이터가 많아야 합니다. 정형이었던 비정형이 됐던 제공되는 데이터가 많아지 활용도가 높아질 거고요. 세 번째로는 사용자의 데이터 마인드셋이 높아져야 됩니다. ... 데이터를 주면 어떻게 활용해야 되고 뭐 이렇게 하는 그런 마인드 셋도 좀 필요하다고 보여 지고요.

3. 선제적으로 탐색적 의미에서 정형을 해 볼 필요가 있고 그 다음에 비정형데이터들을 긁어서 이 선제적으로 탐색했던 내용에 대한 검정이 이루어져야된다는 생각이..네.

7. 그 관계형 데이터베이스와 스프레드시트를 활용한 방법이 좀 저 필요하지 않을까 정형데이터에서는 그 다음에 비정형에서는 뭐 이미 알려진 내용들 SNS 트위터 같은 소셜미디어

를 통해서 이제 비정형 분석을 통해서 미래예측도를 높이는 형태로 사용 활용도를 좀 높일 수 있겠다.

(2) 이용자 입장에서 편리한 청소년관련 빅데이터 체계 활용 방안

다양한 청소년관련 빅데이터를 이용자 입장에서 편리하게 하기 위한 체계화의 방향에 대하여 기본적으로 온톨로지의 중요성이 강조되었으며 응답자 다수는 이용자에 따른 체계화의 중요성을 언급했다. 일례로 연구자나 정책담당자 그리고 일반 국민의 세 부류를 들었는데 연구자의 경우는 데이터의 양이 중요하고 정책 담당자의 경우에는 선별적 데이터의 제공이 필요하다고 지적했다. 나아가 공유시스템의 필요성도 거론되었다. 일례로 미국기관들의 경우 기관은 기관 협의체가 있어 데이터 관리에 관한 협의체의 결정을 따름으로 공유시스템이 작동된다고 지적하였다.

질문 10. 현재 있는 다양한 청소년관련 빅데이터를 어떻게 체계화 하여 활용할 수 있도록 하면 이용자 입장에서 편리할까요?

■ **이용자에 따른 체계화**

3. 문제는 이게 이용자가 누구냐가 중요한 거 같아요 여기 이용자 입장이라고 하지만 일반인을 얘기하는지 아니면 정책을 만드는 사람을 얘기하는지 뭐 아님 뭐 그런 거에 따라서

사실 이게 굉장히 다른 느낌이 들 것 같거든요... 그니까 보는 사람에 따라서 시각화가 굉장히 중요한 것 같아요 ... 연구자 입장에서는 뭐 사실 전체 분석하는 어떤 데이터를 가져온다 어떻게 전처리한다 분석한다 결과를 낸다. 그 다음에 이거에 대한 해석을

어떻게 누구랑 한다 라는 이런 절차가 필요할 것 같아요. 왜냐면 이 분석 절차라는 것도 다 지 맘대로기 때문에... 그니까 검색어를 치면 번들링이 짝 나오고 체크.. 그럼 그거에 따라서 한줄씩 한줄씩 돼있는데 이걸 가져오면 열어보면 다 딴 거 딴 거 예요. 이걸 텍스트 이걸 숫자 뭐 그럼 내가 또 합쳐야되는 거죠. 사실 자동화 온톨로지가 제일 좋은 것 같아요 저는. 데이터를 수집을 하면 우리가 해야 되는 환경에 맞게 뭔가 이제 결합되게 만들어주는 거죠. 짜야 되는데 그거보다 사실은 엑셀에다가 넣어도 공통의 필드가 있는 상태에서 우리가 데이터 넣으면 편하잖아요. 근데 이 필드 하나가 없어져버리면 그 다음부터 이제 힘들잖아요. 이 위에는 필드만 잘 정리가 되면 그걸 지금 하시는 거고 그럼 그 그거에 맞춰서 여기 데이터만 잘 끼어들어가면 나중에 필요한 데이터는 가져오기가 쉬워지니까.. 온톨로지가 사실 제일 중요한 거 같긴 합니다

6. 이게 이용자도 연구자나 정책담당자나 그리고 정말 일반 국민이냐에 따라서 다를 거고요. 그래서 이 세 가지를 나눠서 분류를 하셔야 될 것 같고 연구자들 같은 경우에는 많으면 많을수록 좋잖아요 데이터가 ... 그런 것들을 선택할 수 있게끔 아니면 뭐 가이드를 할 수 있게끔 그렇게 체계화되면 좋지않을까 싶어요. 정책 담당자 입장에서는 너무 많으니까 그걸 다 볼 필요는 없으니까. ...실제 정책 입안자 같은 경우에는 제가 위기청소년 관련된 지표를 무슨 지표가 있는지는 잘 모르겠지만 어쨌든 그 어떤 성과지표를 딱 그걸 보고 싶어요. 그러면 입안해.. 그 성과지표와 관련되어 있는 데이터들이 있을 거잖아요.

4. 기초 마이크로자료는 mds에서 하고 그거를 자기네들이 집계해서 그냥 이렇게 그 일에 지역별이라든지 연도별 그거는 KOSIS에서 하고 그 다음에 국가 전체 12개월 위주로 하

는 거는 e나라지표에서 하잖아요 그렇게 해서 체계화 하면 그 사실은 연구자 입장에선 더욱 감사하죠

5. 이용자 입장에서는 이 KOSIS 시스템이 제일 쓰기가 좋습니다. KOSIS 시스템은 이제 어 통계청에서 수 많은 데이터들을 다 가져가 가지고 1차 가공처리를 한 거죠 ... 구글 트렌드는 되게 간단해요. 그건 그냥 검색 빈도만 보여주는 거니까 그거는 진짜 큰 돈 안들이고 하는 거고 다음은 감정분석까지도 할 수 있게 해주기 때문에 조금 더 공을 들이는 거 거든요. 그 정도로 해주면 정말 좋죠 ...

9. 결국 누군가는 한 군데에서 모으는 사람이 필요하다. 그게 그걸 어떤 그 에이전시가 누가 돼야 될지는 모르지만 그게 정부가 되든 야당 국정위원회 어떤지는 모르지만 그것이 필요하다 생각해요. 그리고 특히나 청소년이라는 그 대상에 대해서

총괄적으로 모으는 주체가 필요하다고 생각해요

■ 공유 시스템의 필요성

8. 요컨대 정보공유입니다 그래서 데이터를 공유해야되는데 각 기관이 공유할 의무가 없어요. 그러니까 우리기관은 우리 기관끼리 두고 재네기관은 재네기관끼리 두고 아까 제가 예를 들은 중에 왜 미국기관들은 이렇게 그 기관 협의체를 만들어서 데이터 관리에 관한 협의체의 결정에 따르기로 돼있단 말이에요. 그러니까 기관은 60 몇개라도 협의기관 20 몇개에서 결정한 대로 해서 전체가 다 올린단 말입니다. 그니까는 공유 시스템이 돌아가는 거예요 ... 데이터를 많이 생산해도 제대로 생산해야되지만 이것이 공유되는 시스템을 만들어야되는데 그게 과연 가능할 것인가 하는 데 대해서 걱정이 돼요.

■ 기타

7. 데이터 그 수집이 처음 필요할 테고 그다음에 어 빅데이터 플랫폼을 만들어서 그것이 효율적으로 사용할 수 있도록 어.. 서비스를 하는...

10. 근데 제일 좋기로는 이제 결국은 청소년들의 그 뭐 우리가 정형화된 데이터들을 중심으로는 일반적으로 우리가 아까 말했던 애들의 뭐 성취도니 뭐 이런 것들은 잘 볼 수 있을 거예요. 그러나 지금 우리가 알고 싶은 것은 정서적인 문제 뭐 이런 것들이잖아요. 실은 청소년 문제하면 이게 근데 뭐 이렇게는 안타깝지만 자살 예후 문제라든지 또 뭐 비행이라

든지 뭐 이런 것들을 알고 싶는데 그거는 이런 걸로는 불가능 이거든요. 정형데이터라는게 결국 너 이렇게 자소서 쓰듯이 뭘 보고하라는 거잖아요. 이제 그거 그렇기 때문에 그런거야말로 비정형데이터가 필요한데 비정형데이터라는 게 결국은 SNS 같은 것들 어떤 형태로든 쓸 수 있는지 한번 보는 것도 좋겠다 왜냐하면 속마음을 털어놓고 드러내는 장면을 캡처해내잖아요. 자연스러운 상태에서 예컨대 퇴근하고 이제 뭐지 하교하고 누구랑 잘 어울리는지를 지금 알고 있는게 없어요. 지금 조사된 게 없잖아요. 근데 저는 굉장히 중요하다고 봐요.

(3) 청소년 빅데이터 체계 구축 및 활용의 장점

청소년 빅데이터 체계 구축 및 활용의 장점에 대해 응답자의 다수는 데이터 활용시의 효율성 제고를 언급하였다. 특히 데이터 전처리에 필요한 비용과 시간절감 그리고 각 기관마다 보유하고 있는 자료의 특성들을 고려한 데이터 연계가능성의 장점을 지적하며 이를 통해 청소년 문제에 대한 선제적 대응, 예방과 청소년의 잠재력 개발 또 청소년의 편의개발 등 다양한 측면에서의 정책적 활용이 가능하리라고 보았다. 다음으로 청소년 문제에 대한 정확한 진단을 통해 사안 관련 현황

파악과 증거기반 정책수립, 나아가 그러한 일련의 정책이 가져올 효과에 대한 예측 가능성이 가능하리라고 보았다. 그리고 이 모든 장점을 통해 결국 안전하고 건강한 청소년을 길러내기 위한 실질적 개선이 이루어질 수 있다고 진단했다. 그 외 일반적 장점으로 서비스 수혜자 또는 참여자의 성향과 상황에 부합하는 서비스를 제공할 수 있게 되어 서비스 품질 향상이 이루어질 것이라는 지적이 있었다.

질문 12. 청소년 빅데이터 체계 구축 및 활용의 장점은 무엇이라고 생각하십니까?	
<p>■ 효율성 제고</p> <p>1. 저희가 데이터를 활용할 때 훨씬 더 효율적으로 활용할 수 있을 것 같아요. 특히 아까 앞서도 말씀드린 것 처럼 데이터 전처리에 필요한 비용 시간절감도 할 수 있고, 데이터를 연계해서 하기도 상당히 좋을 것 같아요 ... 그래서 기관마다 지금 하고 있는 자료의 특성들을 고려해서 다들 근데 아마 그걸 연계를 하는 데는 조금 개인정보라던지 이런 부분들이 좀 많이 어 걸릴 순 있을 것 같은데 그래서 만약 그런 부분들만 할 수 있다면 상당히 많은 연구들을 더 할</p>	<p>수 있을 것 같고 많은 변인들을 활용할 수 있을 것 같다는 생각이 듭니다.</p> <p>6. 같이 여기 생산적인..지금 어쨌든 정책적으로 다양하게 활용할 수 있잖아요.</p> <p>7. 청소년 문제 예방과 또 청소년의 잠재력 개발이라든지 또 청소년의 그 편익개발이라든지 이런 측면에서 필요하다고 생각이 됩니다.</p> <p>■ 정책 진단과 개선</p> <p>3. 선제적 대응을 미리 할 수 있다.</p>

그걸 위한 데이터 구축이다 라고 생각한다면 굉장히 좋은 점이 될 것 같더라는 거 같고요.

9. 그렇게 이제 해놓고 기존에 정부 부처가 막 각자 알아서 해왔던 노력들을 평가해보면 뭐가 문제지를 우리가 진단할 수 있어요. 앞으로 뭐가 개선돼야 될지 그게 두가지 측면에서 맞아요. 하나는 논리적이다. 그리고 두번째는 그 문제점을 개선할 수 있다. 지적하고 개선할 수 있다는 거죠.

■ 현황 파악과 증거기반 정책수립

2. 증거기반에 정책수립과 추진이 일단 가능하다는 것이고요

4. 이게 정책.. 증거기반? 정책수립 당연히 활용 가능하고요

5. 일단은 음 청소년현황 아까 제가 말씀드렸던 청소년현황이죠. ... 파악하는 데 일단 도움이 당연히 될

것 같고요... 청소년 정책 수립이라든가 청소년 정책 관련된 연구에 엄청나게 큰 도움이 될 거라고 생각합니다.

10. 우리 중요한 미래 세대인 청소년들에 대한 것들을 다 좀 건강하고 의미 있는 것들로 만드는데 도움되지 않을까요. 그렇게 해서 아무래도 이제 객관적인 선생님이 말씀하신대로 evidence 기반으로 그렇죠? 어떤 판단을 낼 거 같고 뭐 안전하고 건강한 청소년 사회 속에서 청소년을 길러내는데 도움 되는 일이라고 보고요.

■ 예측 가능성

2. 두 번째로는 예측이 가능해질 가능성이 높아진다는 거구요.

4. 이제 정책의 효과 성과하고 영향 뭐 이런 거를 ... 예측도 가능하고

8. 그런 자료를 바탕으로 우리가 미래를 예측한다든가 더 도움 필요한

<p>게 뭐라든가 현재 고민하고 있는 거 알 수가 있잖아요. 그니까 당연히 이거는 뭐 정말로 귀하죠.</p> <p>■ 서비스 품질 향상</p> <p>2. 세 번째로는 서비스가 향상될 가능성이 상당히 높습니다 ... 특히 그 아까 예측과 관련된 부분에 있어서 서비스 질이 높아질 뿐만 아니라</p>	<p>그 개별화 서비스가 가능해집니다. 그니까 그 정보수집통로만 정확하다면 어 서비스 이용자 개개인의 데이터가 아마 쌓일거고요. 그게 분석되어져서 그 서비스 그 수혜자의 또는 참여자의 성향에 맞게 상황에 맞게 서비스를 제공할 수 있는 여지가 생기기 때문에 기본적으로 서비스 재고 서비스 품질 향상이 굉장히 큰 장점이라고 볼 수 있죠.</p>
--	--

(4) 청소년 빅데이터 체계 구축 및 활용의 단점

청소년 빅데이터 체계 구축 및 활용의 단점에 대해 응답자들의 다수는 개인정보와 관련한 어려움을 지적했다. 일차적으로 개인정보 유출을 통한 사생활침해 가능성이 우려되었다. 자기 자료나 개인정보가 포함된 자료, 특히 비정형데이터들의 무차별적 접근가능성으로 인해 피해를 입을 수 있다는 것 따라서 개인정보 보호를 위한 확실한 법적 장치의 필요성이 지적되었다. 역으로 개인정보보호의 문제로 인한 정보접근의 어려움이 언급되기도 하였다.

다음으로 데이터 품질관리 문제가 제기되었고 데이터의 신뢰성을 보증할 수 있는 정보인증제도의 필요성이 언급되었다. 또한 데이터 구축과 연계과정 등에 필요한 막대한 비용 문제도 지적되었고 그 외 악의적인 데이터 활용, 데이터의 지속가능성 문제가 거론되었고 마지막으로 청소년 문제와 관련한 빅데이터 활용이 제한적일 수 있다는 우려가 있었다.

질문 13. 청소년 빅데이터 체계 구축 및 활용의 단점은 무엇이라고 생각하십니까?

■ 개인정보 차원의 어려움

1. 그거를 관리하는 차원에서는 좀 어려움이 있을 수도 있을 것 같은데 개인정보 라던지 그런 부분들이 자료를 연계하는 데 있어서..

2. 항상 빅데이터는 우려되는 거는 개인정보 뭐 이런건데.. 특별히 단점이라고 생각하는 건 없는 것 같습니다.

3. 개인정보 그거는 좀 문제인 것 같아요 ... 그니까 그 사람의 어떤 개인적인 정보는 취득할 수가 없으니까 그냥 몇 명이 이런 얘기가 많았다 밖에 할 수가 없는 거죠

6. 개인정보.

7. 이것이 이제 잘못 오용되면 사

생활침해 가능성이 있다는 거 그런 것들.

8. 정보의 문제가 있겠죠 개인정보에 대한 문제 특히 사실 우리 또 우리 문화적으로도 이 개인에 대한 정보를 지키는 부분에 대해서 조금 이렇게 무감각한 부분이 있거든요. 그러니까 뭐 confidential한 문제라든가 이런 거 생각도 하고 윤리적으로 전문적 윤리를 생각합니다만은 그럼에도 불구하고 어쨌든 좀 마음이 혼쾌히 내키고 이르지 않아요. 그니까 자기 자료나 자기에 관계된 정보가 들어간 그런 개인정보가 포함된 자료들이 마구마구 제공된다? 마구마구 누구나 접근할 수 있다? 이렇게 생각한다 그러면 되게 불안하겠죠? 특히나 비정형데이터 같은 경우에 알게 모르게 막 긁을 수 있고 하니까 요즘 뭐 스팸이 문제가 아니잖아요. 뭐 피싱이다

별의 별게 다 생기지 않습니까 그리고 그 뭐 최근에 보도에도 그게 악용되어 가지고 갑자기 막 어떤 사람이 타겟이 된다 이럴 수도 있어요 ... 개인정보에 대한 부분은 확실히 보장되는 법적 장치도 만들어져야 될 것이고

10. 단점이나 문제점은 결국은 이제 그 개인정보 문제로 귀결될 것 같아요 ... 최대한 그런 개인정보 문제를 피해 가면서 우리가 원하는 의미 있는 데이터들을 생산하도록 하는 것들은 필요하다 저는 그렇게 생각해요.

■ 데이터 품질관리 문제

3. 뭐 정형데이터는 이미 우리가 잘 가지고 올 수 있지만 비정형데이터를 가져올 때 그냥 막 긁어오느냐 아니면 문제에 맞는 합리적인 것들만 긁어 갖고 오느냐 이건 또 다른 문제인거 같아요.

4. 품질관리 해야 되는데 그거가 통계청처럼 전문기관이면 모르겠는데 그렇지 않은 기관에서는 그것도 부담이죠.

5. 관리가 쉽지 않을 것 같아요 데이터가 어디 있느냐가 중요한데 데이터를 다 모을 것인지 모으면 진짜 관리가 힘들고요. 아니면 뭐 클라우드 같은 거 하나 만들어가지고 그냥 같이 공용을 쓰게 할 것인지 아니면 데이터는 원래 소재에 놔두고 필요할 때마다 이제 그 가공해서 달라고 할 것인지 이런 또 굉장히 또 이게 쉽지 않은 건데 어쨌든...

8. 진짜가 뭐가 진짜고 뭐가 아닌가 그럼 공공적인 데이터는 엄격히 관리돼야 되는 것이고 ... 그 다양하게 생산되는 정보들 그 중에서도 그 이용자들이 혼동되지 않게 구분할 수 있는 뭐 정보인증제도라 할까? 뭐 이게 진짜다. 이 데이터는 정말로 공식

절차를 통해서 수입된 것이다 하는 것을 알 수 있는 어떤 공신력 있는 시스템이나 어떤 구조가 필요하겠죠. 공신력의 확보가 있다면 누구나 편하게 안심하고 이용할 수 있는 이렇게 돼야 되니까.

■ 비용문제

4. 일단 막대한 비용

5. 비용이 꽤 들 것 같은데 이게 지금 저도 어느 정도 비용이 들진 정확히 모르겠는데요 ... 어떤 비용 이런 부분들이 비용이 꽤 방식에 따라서 다르겠지만 뭔가 구축하려고 지금 한다고 그러면 굉장히 큰 비용이 들 것 같다는 생각이 듭니다.

6. 비용. 행정자료들 같은 경우에는 비용이 그렇게 들지는 않죠 처음에 구축할 때의 시스템 구축비용이 들 뿐이지 물론 만약에 그 BPR ISP를 한다면 그 비용도 초반에만

몇억정도 들 뿐이지 한 번 세팅해놓으면 물론 그렇다하더라도 데이터 연계하는 데 또 비용이 들기는 하지만...

■ 부적절한 활용

4. 부적절한 활용. 악의적으로 데이터를 쓰는 사람이 가끔 있어요. 그래서 이제 .. 자기 주장하기 위해서 데이터를 이렇게 분석하는 거죠.

■ 지속가능성 문제

9. 빅데이터로 접근을.. 저는 이렇게 접근하는 게 단점은 제가 볼 때 큰 문제는 이게 지속가능성이 떨어진다는 거예요. 왜냐면 이게 주체가 불분명하다라는 거 그 문제가 여전히 있고.

■ 활용상 제한

9. 빅데이터에 대한 접근은 저는

<p>그렇게 생각해요. 이게 빅데이터로 할 수 있는 우리가 봐야 되는 청소년이 이만큼 있으면 빅데이터로 할 수 있는 건 매우 제한적이예요. 그래서</p>	<p>나는 이 많은 블랭크가 생길거다. 전통적인 지표를 하려고 하지 않으면. 그게 단점이죠.</p>
---	---

(5) 한국의 상황에서 청소년 빅데이터 체계 구축 및 활용의 한계점

한국의 상황에서 청소년 빅데이터 체계 구축 및 활용의 한계점에 대한 질문에 대해 응답자의 다수는 데이터 확보 문제를 지적했다. 데이터 확보가 어려운 구체적인 이유로서는 먼저 개인정보 보호 정책으로 인해 자유로운 정보활용의 한계가 있다고 답했다. 둘째, 데이터공개 시점의 문제점을 지적하였는데 자료 클리닝 작업 등으로 적시에 공개되지 않는 경우가 있으며 동일한 맥락에서 정리가 되지 않은 데이터가 있을 수 있다고 응답했다. 나아가 데이터 확보가 어려운 이유로 셋째, 수집구조의 문제를 지적했다. 플랫폼이 있어 데이터가 수집되고 데이터에 근거한 서비스가 가능한데 그런 수집구조가 없다는 것이다. 특히 중요 데이터들의 경우 다 협조를 받아야 한다는 어려움을 언급하였다. 그리고 많은 경우 데이터를 공개안하거나 하더라도 활용성이 적은 데이터인 경우가 적지 않다고 지적하였다. 결국 전체적으로 보면 컨트롤타워의 필요성이 강조된 것이라고 볼 수 있다. 이 외 데이터 확보보다 오용문제를 지적한 응답도 있었다. 빅브라더같이 개인의 일상이 국가통제나 기재로 오용될 수 있다는 것 따라서 빅데이터의 구축과 함께 관리의 중요성을 강조하였다.

질문 14. 한국의 상황에서 청소년 빅데이터 체계 구축 및 활용의 한계점은 무엇이라고 생각하십니까?

■ 데이터 확보 문제

1. 자유로운 정보활용의 한계

1. 개인정보 그 보호법 때문에 자료를 이렇게 연계하는 작업이 상당히 어렵더라고요. 왜냐면 저희가 주민등록번호를 사용할 수가 없으니까. 그런 부분들이 좀 어렵고.

3. 아까 얘기했던 개인정보법 그게 좀 제일 어려운 부분이 아닐까.

4. 저는 지나친 개인정보 보호 정책? 아 뭐 하나 데이터 그 미국의 센서스뷰로 인구자료 데이터 제가 써봤거든요. 그 거는 이메일로 신청을 해요. 다음날 바로 데이터가 이메일로 날라와요. 그냥줘요. 거긴 소득도 다 들어가 있어요

2. 데이터공개 시점의 문제

1. 자료가 공개되는 시점이 또 중요하다라는 생각이 드는데 ... 저는 KCPIS 데이터는 상당히 빨리 빨리 오픈이 되는 것 같아요. 근데 그 외에 이제 다른 기관이나 아니면 뭐 교육청 같은 경우에 전 조금 자료를 오픈하는 데 있어서 아마 자료 클리닝 작업이 조금 오래 걸려서 그러신 거 같더라고요... 그리고 정리도 항상 잘 돼 있는 편 KCPIS 자료가..그에 반해서 이제 그런 정리가 좀 안돼 있는 데이터들도 있고 뭐 그래서 활용하는데 그런 게 좀 많이 어려움이 있었었던 것 같아요.

3. 수집구조의 문제

2. 수집구조가 없는 거예요. 빅데이터를 어떻게 모아내고 찾아내갈건지 하는 건데 ... 청소년층 같은게 플

랫폼이 돼서 거기서 어 청소년들의 데이터가 수집되어지고 그거를 가지고 그 청소년들에게 좀 더 필요한 서비스라든가 청소년 생활에 맞거나 실 생활에 맞는 또는 청소년들의 요구에 맞는 그런 서비스를 만들어낼 서비스와 정책을 만들어 낼 수 있을텐데 그런 수집 구조가 없다라는게 그렇다고 강압적으로 9세부터 20세 사이 청소년들한테 무슨 앱을 깔게 할 수도 없는 거잖아요. 네 그러니까 그런 수집 구조를 그 어떻게 만들건가 하는게.

5. 데이터 확보가 어렵다 빅데이터를 얘기할 때 많은 사람들이 걱정한 게..에 대해서 데이터가 있냐? 응? 아 데이터야 있죠 근데 데이터를 갖다가 쓸 수가 있느냐 빅데이터를 그게 지금 가장 어려운 거라고 보는 것 같아요 ... 중요한 데이터들은 다 협조를 받아야 되거든요. SNS 기업체라든가 또는 뭐 네 검색 기업체라든가 카드사라든가 그렇게 해야 되니까 ... 그러면 애네들이 제일 많이 들어

가는 데가 뭐냐 결국은 유튜브하고 SNS 페이스북 뭐 인스타그램이 제일 많이 들어가고 그러니까 인스타그램은 그러면 협조를 얻어야 되는데 거기가 협조해주겠냐 그러면 유튜브도 마찬가지로 그래서 하.. 이게 좀 저도 지금 좀 고민거리입니다. 그래서 그런 거를 어떻게 확보하냐가 일단 제일 큰 장벽인 거 같더라고요 저는.

4. 컨트롤타워의 부재

7. 컨트롤타워가 없다는 게 크죠

9. 문제는 청소년 정책은 부처에서 각 나눠서 다 개인별로 하고 있으니까 이것 이거 그렇게 말하면 종합적인 보고서가 없어요. 그러니까 이게 문제죠 그래서 저는 지금 현재로서는 제일 중요한거는 저는 종합보고서가 나오는데.. 아동종합보고서 청소년종합보고서가 나와야 된다고 생각해요.

5. 데이터 미공개 또는 부분공개

6. 데이터 공개안함

8. 예컨대 기관이기주의라는 말도 있는진 모르겠습니다만 기관 나름대로의 상황이 있겠죠? 그런 것이 알려질 수도 있겠죠? 이 데이터의 공유가 된다는가 알려지게 되면. 그래서 공개가 필요하지만 공개를 할 수 없는 구체적인 자료같은 건 아마 공개 안 할 수 있을 거예요 그러니까 어그러니까는 외형적으로는 공개하는데 실제적으로는 정말 필요한 자료는 공개되지 않고 있을 수 있다 ... 그러니까 그리고 형식적인 관리? 관리한다 뭐 정보 빅데이터 센터를 만들라 하면 그냥 만들면 되는 거예요 근데 그게 과연 그게 제대로 기능하는가 제대로 된 역할을 하는 가는 그 제대로라는 말에는 생각이 다를 수 있겠죠 근데 그게 기관마다 다 만들라 하면 막 만듭니다 근데 들여다보면 별로 내용이 없는거예요. 진짜 그게 믿을만 할 것인가 진짜 꼭 필요한 자료는 못찾는

다 이런 상황도 생기지 않을까.

■ 정보의 오용문제

10. 그 왜 조지 오웰 소설에 나오는 것처럼 빅브라더 같은 상황 곤란하다라는게 저는 기본적인 생각이고요. 특히 코로나 때문에 더 느껴지는데 어쨌든 여기 오실 때도 뭐 다 쓰고 오시잖아요 ... 이게 일상 속에서 삶 속에서 국가통제나 이런 게 기재로서 활용되는 것은 적절하지 않.. 까딱 잘못하면 이럴 수 있어요 ... 그래서 그게 핵심이었어요. 그게 뭐 빅데이터를 잘 만드는 것도 중요하지만 그런 일들이 관리하는 일이 좀 돼야될 것 같고 ... 또 그 우리가 모으려는 데이터들이 지금 말씀하신 부분처럼 청소년들이 관심 있는 데이터를 모으고 있는 것이냐 아니면 우리 관심 있는 걸 모으는 것인가.

(6) 빅데이터를 잘 활용하기 위한 개선방안

빅데이터를 잘 활용하기 위한 개선점에 대해 응답자들은 일차적으로 컨트롤 타워의 필요성이라고 응답했다. 정형데이터를 표준화하여 정리하고 또 연계 사례를 공유할 수 있도록 해주는 컨트롤타워가 필요하다는 것이다. 동일한 맥락에서 데이터 확보, 공유 및 서비스하는 시스템의 필요성에 대한 응답이 있었는데 이는 홍보와 체계적 관리, 그리고 소비자 중심의 자료 가공 및 제공의 3가지로 정리한 응답과 맥을 같이한다.

결국 정리하면 컨트롤타워가 있어 데이터 확보, 공유, 품질관리 및 서비스하는 시스템이 필요하다는 것으로 정리할 수 있다. 그 외 빅데이터를 잘 활용하기 위해서는 데이터 개방, 역량 있는 분석가의 양산, 담당공무원의 추진력, 그리고 수직구조 개선이 필요하다는 응답이 있었다.

질문 15 빅데이터를 잘 활용할 수 있도록 하기 위해 어떠한 부분이 개선되면 될까요?	
<p>■ 컨트롤 타워 필요성</p> <p>1. 이젠 이제 정형데이터에 관련된 건데 그거를 이제 컨트롤 타워가 만들어 드린다면 그 위에서 어떤 데이터를 정리하는 그런 것들에 대해서 하나의 룰을 만들어 주시면 밑에 있는 기관이나 아니면 뭐 교육청에서 그거에 맞춰서 데이터를 정리한다던지 이렇게 표준화 같은 작업들을 하면 정형데이터는 훨씬 좀 활용하기</p>	<p>좋을 것 같고요.</p> <p>7. 빅데이터 활용을 지원하기 위한 연계 사례를 공유할 수 있는 그런 컨트롤타워의 구축이 필요하다고 생각됩니다.</p> <p>■ 데이터 서비스 시스템</p> <p>4. 간단합니다 홍보 그다음에 체계적인 관리 그다음에 빅데이터기 때문</p>

에 소비자 중심의 자료의 가공 및 가공이 3가지만 되면 되는데 쉽지가 않죠. 예산하고 인력하고 그 다음에 또 나중에 행정적으로도요. 이게 뭐 이게 이거는 이 데이터는 안되고 이런 거 나중에 걸려요.

5. 데이터 소스를 원소스가 가지고 있고 우리는 유저인데 중간에 누군가가 이걸 갖다가 우리가 데이터를 원데이터를 갖고 있다는 해도 전 처리까지 해가면서 다 해서 하기에 저는 어렵다고 보고 누군가가 중간에서 역할을 해주는 어떤 이걸 플랫폼이라고 표현해야 될지 모르겠는데 ... 그러니까 이거는 약간 데이터 서비스하기 위한 어떤 어떤 이런 그런 회사 그래서 저는 데이터 확보 공유하고 서비스하는 어떤 시스템이 좀 마련되는 게 중요하지 않나 생각이 들었습니다.

6. 개방되어 있는 데이터들의 품질 관리가 제대로 안되고 있어요. 품질

관리부터 해야되지 않을까 싶어요.

7. 빅데이터가 이게 퀄리티가 있는지에 대한 문제가 좀 중요한 거 같고 신뢰성...

8. 정보의 신뢰성 그리고 정보 관리의 체계성 뭐 이루 말할 수 없습니다.

■ 분석가의 역량

3. 분석가의 역량이 참 중요하다고 생각해요. 왜냐면 머신러닝을 합니다 라고 하지만 머신러닝 방법론 아는 사람이 그렇게 많다고 저는 생각하진 않거든요 그니까 텍스트마이닝을 합니다도 마찬가지고. 그니까 텍스트마이닝이 뭔지는 알아요. 분석하는 방법이 뭔지도 알지만 실제 분석을 할 수 있는 사람은 거의 없어요.... 분석가 양산은 상당히 필요하더라는 거죠.

<p>■ 담당공무원의 추진력</p> <p>6. 담당공무원의 추진력을 말씀 드리잖아요 공무원들도 어 이거 빅데이터 정말 필요해 빅데이터 분석해 가지고 우리가 연구하는 게 필요해라는 게 필요해요</p> <p>■ 데이터 개방</p> <p>6. 데이터 개방의 부분이 개선되어야 하고,</p> <p>8. 역시 그 정보 접근성 이겠죠 접근성이고</p> <p>9. 그 다음에 자료 접근 가능성이 떨어져도 뭐 사실은 나이스 이런 좋은 정보가 있어도 우리가 알 수 없잖아요. 접근이 안되잖아요. 그런데 접근해야 된다. 근데 그거는 사실 박사</p>	<p>님 우리의 노력이 아니고 이걸 제도적인 문제라서 시간이 걸리고 그런 거죠.</p> <p>10. 활용되려면 뭐가 핵심일까 데이터의 오픈이라고 봐요. 근데 조금 청소년 정책연구원도 보니까 데이터를 몇 개를 했더라고요. 가지고 있더라고요 이제 매년 패널? 내가 그거 좀 써보려 했더니 뭐는 몇년 전거만 주고 또 그 뒤에꺼 안주고 뭐 이렇게 되어있더라고요.</p> <p>■ 수직구조 개선되어야</p> <p>2. 수직구조가 일단은 개선되어야 된다고 봅니다 그래서 아까 말씀.. 14번 15번이 아.. 그 한계점과 개선점이 저는 기본적으로 수직구조에 있다고 보여집니다.</p>
---	--

(7) 청소년정책에서 빅데이터 활용방안

청소년정책에 빅데이터를 어떻게 활용할 수 있을지에 대한 응답은 크게 두 가지였다. 첫째는 증거 기반 정책 수립이고 둘째는 실효성 있는 정책 개발이다. 전자와

관련하여 빅 데이터는 실증자료를 활용한 정책 결정은 물론 정책에 대한 사전 사후 평가를 가능하게 해줄 것으로 평가했다. 또 위기청소년이나 청소년의 정신건강 등 구체적인 사안들에 대하여 실효성 있는 정책, 즉 정책을 위한 정책이 아닌 실무적 시사점을 도출할 수 있는 정책개발에 기여할 것으로 평가했다. 그 외 기존의 연구방식이 주지 못하는 새로운 정보를 도출할 수 있고 청소년과 청소년 문제에 대한 관심을 유발할 수도 있으리라고 내다봤다.

<p>질문 16. 청소년정책에 빅데이터를 어떻게 활용할 수 있다고 생각하시는지요?</p>	
<p>■ 증거 기반 정책 수립</p> <p>1. 네. 이건 아까 박사님께서도 말씀 주신 것처럼 실증자료를 활용한 정책 결정 또 증거기반 교육실천 이런 게 이제 미국에서도 많이 강조하고 있었고 관련해서 지금 미국에서도 정책을 위해서 그런 데이터를 이제 빅데이터 작업을 해놨더라고요.</p> <p>2. 아까도 말씀드렸듯이 일단 증거에 기반한 정책을 세울 수 있는 것이기 때문에 그런 면에서 도움이 될 거고요.</p> <p>4. 자료를 통해서 근거를 도출하거나 정책을 사전 사후적으로 평가하거</p>	<p>나 이런 거에 충분히 쓸 수가 있죠.</p> <p>5. 그러면 이 정책이 맞는지를 갖다가 따져보거나 이럴 때 도움이 되겠죠. 그래서 정책이라고 하는 게 꼭 정책을 개발하는 것도 있지만 정책을 개발하기 전에 여러 가지 해야 될 사전 작업들이 있잖아요. 뭐 수요자의 견 수렴도 있고 이슈를 파악하고 동향을 파악하고 이런 데 이제 이 빅데이터들이 좀 도움이 되겠죠.</p> <p>■ 실효성 있는 정책 개발</p> <p>3. 어.. 뭐랄까 기존에 그냥 사실 어떤.. 아 근데 정책적인 정책 말고 실무적인 어떤 시사점을 만들어 낼</p>

수 있는 정책을 개발할 수 있다라는 점이 장점이 될 거라고 생각합니다.

6. 그 제일 일단 요구 니즈가 많은 게 위기청소년 아니예요? 제가 생각했을 때. 그리고 일단 정신건강. 같은 개념일 것 같아요. 저는 정신건강 쪽으로 그냥 이야기를 하는 게 좋을 것 같아요. 왜냐면 청소년 정책이라고 하면 그 정신건강 관련된 조사도 하잖아요 ... 어쨌든 이런 요소들이 잘 갖추어져 있으면 이렇게 정신건강에 문제가 없고 아주 건강한 아이들이 이제 뭐..더라 그런 부분들이 보여지면 개를 벤치마킹할 수 있는 뭔가가 만들어 질 수 있지 않을까요?

8. 그니까 이 데이터가 청소년에 관한 데이터도 중요하죠 그러나 청소년 정책에 관계된 소위 말해 청소년 업무가 어떻게 가고 있는가에 대한 걸 파악해서 청소년 정책의 수준 청소년 정책에 제안할 방향이 과연 선진국의 상황에 우리가 얼마나 접근했는가 아까 제가 국제 기준 비교하는

그 지표를 얘기했는데 그렇게 맞춰져야 한다는 얘기에요.

■ 기타

7. 16번은 또 앞의 거하고 연결이 그대로 되는 거라서..

9. 저는 그 이거는 답을 못하겠더라고요. 솔직히 말씀드리면. 이게 아까 그 제가 아까 자살 연구에서 말씀드렸듯이 그거는 그렇게 하면 명확하게 쓸 수 있어요 그 완전히 새로운 방식의 새로운 정보를 다 도출할 수 있어요.

10. 왜 패널데이터 가지고 왜 무슨 연구대회 같은 거 하잖아요 결국은 데이터를 오픈하는 일도 중요하고 그것 데이터가 활용되도록 다양한 연구자들 관심 있는 사람들을 계속 모아서 하는게 나는 청소년 그 팀 연구팀에서 하실 일이다 자꾸 오픈해서 관심 갖게 막 계속 하고 그런 무슨 학술대회 같은 게 저는 좋은 샘플 같아요.

8) 청소년 대상 빅데이터 체계 구축 방안

(1) 청소년 대상 빅데이터 체계 구축을 위해 선결 사항

청소년 대상 빅 데이터 체계 구축을 위해 선결되어야 할 사항으로 응답자들의 다수는 자료의 통일적 체계구축이 필요하다고 지적하였다. 다양한 연구기관이나 또는 교육청 등의 자료들을 체계적으로 관리, 활용하기 위해 온톨로지가 먼저 필요하다는 것이다. 다시 말해 빅 데이터 분석과 운영관리 체계 수립을 통한 빅 데이터 분석체계와 빅 데이터 통합관리 방안 수립, 그리고 빅 데이터 인프라 환경 분석과 도입계획을 수립하기 위한 시스템 환경 등이 필요하다고 보았다. 다음으로 응답자들은 자료 관리의 측면에서 질관리 프로그램을 위한 거버넌스의 필요성을 제기하였는데, 이에 상응하는 응답으로 구성원들의 마인드셋이 디지털 친화적 내지는 데이터 친화적으로 바뀌어 패러다임 전환을 통한 수직구조 개선이 요구된다는 입장, 그리고 필요성 및 활용성에 대한 가치 인식 공유, 빅데이터 구축과 데이터 수집에 있어 청소년과 학부모에 대한 인식 고려가 필요하다는 주장이 있었다. 그 외 청소년만의 이슈가 될 요소들에 대한 사전 설정, 예산 및 인력 확보 그리고 데이터 개방에 대한 지적이 있었다.

질문 17. 청소년 대상 빅데이터 체계 구축을 위해 선결되어야 하는 사항은 무엇이라고 생각하시는지요?

■ 빅데이터 분석과 운영관리 체계 수립

1. 그래서 그거를 제일 이제 다양한 연구기관이나 또는 교육청에서

모아 온 정형자료 같은 것들을 하나로 모아서 뭐 활용할 수 있는 방안을 좀 논의하면 좋겠다는 생각이 들었고 그래서..

3. 사실 그 체계가 제일 중요할 수밖에 없긴 해요 온톨로지 자체가 중요한데 온톨로지를 구상하기 위한 기본적인 그 논리는 굉장히 중요한 것 같아요. 그냥 막 중구난방 이런 게 아니라 우리의 중요한 데이터는 애는 필수적으로 뭐가 필요하니까 온톨로지에 이게 들어가야 된다 이게 다 정해져야 된다는 거죠. 사실은 그게 선결되어야 된다고 생각하고.

7. 첫 번째는 빅데이터 분석과 운영관리 체계 수립이 중요하다고 생각이 되고요. 이를 통해 이 세부 내용으로는 빅데이터 분석체계와 빅데이터 통합관리 방안이 수립돼야 되고 두 번째는 빅데이터 인프라 환경 분석과 도입계획을 수립해야 되는데 여기서는 빅데이터 분석을 위한 시스템 환경 또 보유데이터 현황을 분석하는 것과 빅데이터 시스템 구축 계획 수립 이런 게 필요한 거 같고 세 번째는 빅데이터 인프라 구축과 기본분석기능을 개발하는 건데 빅데

이터 분석처리 시스템과 인프라 설계 플랫폼 구축 분석 시각화 개념 개발이라든지 그다음에 수집 저장 처리 분석 시각화 자극 처리를 검증하는 내용들이 필요하고 그다음에 빅데이터 분석을 위한 데이터를 확보해야 되는데 그거는 내외부 데이터 수집을 위한 기존 시스템 연계 개발과 데이터수집 그리고 이제 수집데이터를 처리할 수집한 데이터를 처리 개발 수행하는 능력 이런 것이 필요할 것 같습니다.

9. 저는 뭐 아까 말한 부처 간의 협업하고 자료 접근성을 어떻게든 높이는 게 그게.. 그게 그리고 그것에 대한 프로토콜이 있어야 돼요. 어떤 절차를 어떻게 밟아서 어떻게 하고 그 다음에 어떤 마스킹문제는 어떻게 하고 그런 것들에 대해서 전체적인 그 말하면 모듈이 딱 있어 가지고. 근데 그거는 제가 보기엔 통계청이나 이런데서 해야돼요

10. DB를 계속 그니까 확장해서
담아낼 수 있는 그릇을 처음에 설계
해놓으시면 좋을 것 같아요.

■ 가버넌스

4. governance요 누가 관리 할
거나 지금 이제 청소년 말이 주제는
청소년이지만 박사님 이 데이터가
다른 엉뚱한 부처에서 올 수도 있잖
아요. 그래서 임대주택 거기 청소년
들어가고 있고 그럼 그거는 이쪽 이쪽
에서 오고 다른 뭐 청소년 자료 뭐
다른 부처 오고 그러잖아요. 그럼 그
자료에 대한 권리 관리할 수 있는 권
리 누가 가질거나 권한을.

8. 네 그러니까 어디까지는 뭔가
이렇게 좀 그런 특히 퀄리티 관리 프
로그램 퀄리티 관리는 반드시 공적
으로 해야 되잖아요. 이런 거 까지도
그냥 맡겨놓으면 안되지 않습니까.
그래서 형식적으로도 안되는 부분은
우리가 체계적으로 중앙에서 규정화

를 하거나 제도적으로 해야되고 이
런 생각이 듭니다.

■ 인식구조의 패러다임 전환

2. 수직구조 개선 그 다음에 그 정
책적 전환 그 다음에 그 이 안의 구
성원들의 마인드셋이 디지털 친화적
내지는 데이터 친화적으로 바뀌어야
된다 라는 거고 그건 금방 바뀔 것
같은 생각을 들어요.

5. 필요성 그리고 이 활용성에 대
한 어떤 가치에 대한 인식이 공유되
는 게 일단 제일 중요하지 않나 라고
생각이 들고요.

10. 이제는 빅데이터 구축과 데이
터 수집에 청소년과 학부모에 대한
어떤 인식도 좀 고려했으면...

■ 청소년 이슈 사전 설정

3. 제일 중요한 건 일단 그런 거

<p>같아요. 그 성인하고 구별해서 청소년만의 이슈가 될만한 요소들을 미리 확인해놓는거.</p> <p>■ 예산 및 인력 확보</p> <p>5. 그리고 예산과 인력이 확보되면 어 추진해 나갈 수 있지 않을까?</p>	<p>■ 데이터 개방</p> <p>6. 데이터 개방. 복지부라든지 그런 쪽은 어쨌든 데이터 사회보장정보원에서도 데이터 제공할려고 노력중이에요. 그리고 식약처도 공공 데이터 개방전략 세우고 있잖아요. 근데 청소년 쪽에서 그런 연구 못봤어요...</p>
---	---

(2) 빅데이터 체계 우선 구축 주제

빅데이터 체계구축에 있어 우선적으로 고려해야할 주제에 대해 응답자 다수는 청소년의 삶과 행복과 관련한 것이어야 한다고 답하였다. 구체적으로는 청소년들의 시간활용, 가정 내 아동학대 및 폭력, 그리고 중독이나 자기조절 문제 등의 소극적 영역과 청소년의 신체적, 정서적 건강 및 안전 분야, 인지적 영역의 학업성취도 등 적극적 영역의 주제들이 언급되었는데 특히 자율성의 주제가 강조되었다. 기타 실천적 주제로 봉사활동이나 사회적 기업 등 시민적 관여의 문제도 거론되었다. 그리고 그 외의 응답은 주제가 아닌 방법론과 관련된 것이었다. 즉 정보파악 총괄과 빅 데이터 팀의 구성, 데이터 확인, 정형 비정형의 매칭 되는 요소들을 통한 주제설정 및 온톨로지 설계 및 활용, 그 과정에서 목적성에 근거한 분류체계, 또는 표준코드 수립 등이 거론되었다.

질문 18. 다양한 빅데이터 체계를 순차적으로 구성할 경우 우선적으로 먼저 구축해야 하는 주제는 무엇이라고 생각하시는지요?(혹은 가장 중요하다고 생각하는 사항은 무엇인지요?)

■ **삶과 행복**

1. 저는 약간 정의적 행동 특성에 조금 더 관심이 요즘 많이 많아진 것 같아요. 그래서 아까 말씀드렸던 뭐 요즘 사회적으로 문제되고 있는 뭐 중독이라던지 자기조절이라던지 뭐 이런 부분들이 의외로 패널자료에서는 많이 활용하기가 어려웠었어요 ... 기존에 있는 정형데이터에서는 인지적영역, 그 학업성취도와 관련된 거는 오히려 많이 있었으니까 약간 정서적인 부분이 좀 정의적인 부분들을 많이 다룰 수 있으면 좋을 것 같다는 생각이 좀 들었어요.

4. 행복. 청소년의 행복. 그 다음에 이제 청소년들의 시간활용. 그 다음에 가정 내 부모에 의한 아동학대 및 폭력 이런 게 좀 시급하지 않나 생각합니다.

8. 가장 중요한 거는 청소년의 삶에 관한 것인데 ... 우리는 늘 문제로 접근하거든요. 청소년들이 어떤 사고를 쳤나 학교폭력을 얼마나 당했나 이렇게 접근하잖아요. 청소년들이 건강한가 생각해서 정서에서 신체적인 상황에서 가족관계에서 얼마나 건강한가 생각과 모든 거 이렇게 조금 프레임을 다르게 접근하는 이런 거 제가 한번 좀 중요하다 생각하고요. 해서 하여튼 그 문제 있죠. 건강과 관계해서 우리 안전 분야 생각해야 해요. 안전을 무슨 뭐 수련활동 가서 청소년 수련활동 가서 사고난 그거만 생각하는데 곳곳에 ... 일상 생활에서도 있습니다...

9. 네 저는 아까 그 말들 그 행복함이나 성격 그게 제가 보기에는 우리나라에서 청소년 문제에.. 청소년 이슈에서 제일 큰 문제가 뭐냐 나는 내

맘대로 못하는 거라 생각해요. 애들이 자기 맘대로 못하는 거 그 자율성을 높이는 방향으로 우리가 가야된다는 거예요.

■ 기타 (실천적 주제)

8. 시민적인 관여? 어 뭐 봉사활동이라든가 사회적인 기업? 이런 데 대한 것도 우리 좀 알아야 되지 않겠나 생각해요.

■ 기타 (방법론)

3. 그 저기 정형데이터 정형 비정형의 매칭되는 요소들 그리고 해결해야될 주제들 그걸 정하는 게 먼저다. 그 다음에 온톨로지 설계가 되고 온톨로지에 데이터가 들어가고 그럼 이제 그 데이터를 끄집어 와서 분석

을 하고 이런 순서 일 것 같아요.

5. 네. 일단 저는 데이터를 갖다가 확인. 어떤 데이터가 있는지가 일단 제일 아는 게 중요할 것 같습니다.

6. 분류체계 선정 분류체계.. 그리고 목적성이 있어야 될 것 같은데요.

7. 기반조성 단계로서 그 정보파악 총괄과 빅데이터 팀이 만들어져서 대내외 협의체를 구성하기도 하고 참여하는 이런 것이 필요하다고 생각됩니다.

10. 그 과정에서 혹시 그게 불가능한지는 모르겠지만 일반적 코드의 표준화라고 우리는 그냥 통상 얘기하는데 표준코드 같은 걸 만드는 작업을 한 번 하실 필요는 있어요.

(3) 청소년 관련 빅데이터 관할 컨트롤타워 필요성 및 역할

컨트롤 타워의 필요성과 관련하여 응답자들은 모두 필요성에 공감하고 있었다. 그 이유로는 무엇보다 분산되어 있는 데이터들의 상호 개방과 연계를 통해 효율성을 높이는 것이다. 다만 실질적인 관심보다는 과시적 차원의 관심에 대해 염려하는 의견이 있었다. 다음 컨트롤 타워를 어디에 둘 것인가 하는 문제는 청소년 전문 유일한 국책연구기관인 한국청소년정책연구원이 맡아야 한다는 다수 의견과 현재 모든 공공 자료를 관리하는 통계청에서 컨트롤타워로 역할을 하는 것이 데이터 활용상 홍보 강화 그리고 품질 관리 측면에서 효율성을 제고할 수 있다는 입장, 그리고 기타 핵심기관의 필요성만 강조한 견해가 있었다.

질문 11. 청소년 빅 데이터 체계 구축 및 활용을 하기 위한 대안으로 청소년관련 빅데이터를 관할하는 컨트롤 타워를 만들어서 자료의 활용성을 높이는 것에 대해 어떻게 생각하십니까?

11-1) 컨트롤 타워가 필요하다면 그 이유가 무엇이고, 어떤 부처와 기관에서 컨트롤 타워 역할을 해야 한다고 보십니까?

11-2) 컨트롤 타워가 필요하지 않다면 그 이유가 무엇입니까?

<p>■ 컨트롤 타워의 필요성 이유</p>	
<p>1. 효율성</p> <p>1. 당연 필요.. 필요 하다는 생각은 들어요. 왜냐면 지금 아까 말씀드린 것처럼 교육청에서는 자기 시도에 맞는 그런 정책에 포커스 뒤서 지금 이미 하고 있고 자료를 모으고 있고 뭐 다른</p>	<p>각 기관 마다 이제 설문을 하고 있고 근데 그런 자료를 다 연계할 수 있다면 훨씬 더 효율적인 정책을 할 수 있을 것 같아요. ...</p> <p>3. 애들에 대해서 사실 다들 관리감독은 하고 있지만 우리가 선제적으로 하는거냐의 관점에서의 컨트롤타워라면 필요하다고 생각을 하거든요</p>

8. 어쨌든 데이터라는 말이 주는 것은 연구자료 아닙니까. 제가 볼때. 그래서 연구 중추적인 어떤 연구기관에서 이거는 좀 관리할 필요가 있다는 생각이 들고 저는 개인적으로 어.. 이 청소년관련 데이터가 너무 분산돼 있어요. 활동 관련 데이터 네 그다음에 상담관련 데이터. 연구관련데이터 하여튼 음..

9. 저는 필요 있다고 생각해요. 지금 이거 이대로 놔두서는 저는 절대 안될거라 생각해요 ...모여서 같이 해야 되죠. 그걸 그러면 어디서 누가.. 그러니까 아무도 안하려고 하는 거예요. 그러니까 안된 거고 많이 비어있죠.

10. 컨트롤타워는 그렇게 자연스럽게 존재 이유가 없는 것처럼 실제로 작동되도록 만드는 의미에서 저는 필요하다고 봐요 ... 지금은 컨트롤타워가 부처 간에 장벽을 낮추고 협업을 이끌어내는 일이 됐으면 좋겠다는

거예요. 왜냐하면 이 데이터를 생산하는 입장에서 보면요. 이 기관에서 달라 해서 또 주고 저 기관 이렇게 할 수는 없잖아요. 그러니까 어떤 특정 장면이 캡처가 되면 그것이 다 모든 사람이 공유될 수 있도록 하는 것... 그거를 조금 이렇게 뭐라 그럴까 전략적 제휴인지 협업 분업인지 해서 서로 아 ...그러니까 모든 것을 다한다고 하지 말자 이거죠.

2. 문제점

10. 근데 지금 빅데이터에 컨트롤 타워는 어떤 의미에서는 품잡기 위한 거 같은 느낌도 사실은 미안하지만 들어요. 왜 막 빅데이터 합니다. 보니까 우리도 한 건 한다. 야. 우리가 뭐 한다. 이제 그런 의미에선..

■ 담당 부처와 기관

1. 제 3의 기관

1. 국무총리실은 제가 생각을 하지 못했었는데 그럼 국무총리실 밑에 산하 기관을 따로 두는 건가요. 아니면은 뭐.. 그것도 좋을 것 같은데요. 제 생각에는 네. 그거는 제가 정보가 없어가지고 따로 생각을 못 했는데 별도의 기관을 약간 위에 두는 것이 가장 데이터를 모으기에는 효율적이지 않을까 싶어요.

■ 한국청소년정책연구원

2. 전적으로 동의하고요 데이터를 뭐 자체적으로 어떻게 갖고 갈건지에 대한 정책을 갖고 있는 건 굉장히 중요한 부분이기 때문에 어 컨트롤타워가 필요하다고 보여지고요 ...그 한국판 뉴딜에 데이터뉴딜 같은 경우는 어 일단 부처가 중심되고 기재부가 중심입니다. 그래서 그 부처가 중심이 되고 관련 기관들이 전문성을 보태는 방식이거든요 그런 면에서는 ...그니까 저는 그 그런 역할이 정책연구원에서 하셔야 되는 역할이라고 보

고 잘 잡으신거 같다는 말씀을 드립니다.

3. 좋다고 생각해요 어차피 그 뭐랄까요 그리고 어차피 저희가 그 뭐 청소년연구원은 사실 딱 이쪽에 딱 맞춰서 이렇게 특화된 기관이잖아요 그냥 사실 이 기관에서도 할 수 있는 문제라고도 생각이 들기도 하고 사실은 컨트롤할 필요는 있겠죠 이게..부처는 뭐 사실 뭐 이쪽에서 하는 게 맞지 않나요? ... 연구원이 청소년에 대한 거니까 아무래도 그 쪽 연구를 제일 많이 해오셔서 제일 많이 아시잖아요. 그런 분들이 문제를 도출한 다음에 분석하게끔 만들면 좀 틀릴거라고 저는 생각해요. 그게 예를 들어서 이걸 갑자기 행안부에서 하시거나 아니면 뭐 과기부에서 하거나 과기부에선 당연히 이 빨리 분석하는 방법에 집중하시게 될거예요. 행정안전부는 뭐.. 폭력, 자살 이런 쪽은 굉장히 신경 많이 쓰시겠지만 또 다른 세세한 부분은 또 관심이 없게 될 수도 있어

요. 그렇다보니까 이 특수집단이잖아요. 청소년이라는건. 특수집단에 대해서 전담하는 기관이 있으니까 거기서 하는 게 맞다고 저는 생각이 들어요

5. 당연히 컨트롤타워 필요하다고 봅니다. 그게 없으면 추진하기가 지금 어렵다고 보고요. 그 뭐 청소년 정책연구원이 해도 좋을 것 같고 다만 이제 청소년에 한정해서는 효용성이 조금 제약될 수 있다고 생각이 아까도 말씀드렸지만 그런 게 좀 드는 거예요. 그래서 좀 이제 나중에 이제 우리 데이터를 갖다가 어떻게 구축해야 될지 조금 더 고민해가면서 해봐야 되는데 청소년, 청년 그러니까 대상을 보면 청소년과 청년이 연결되고 그 다음에 이제 청소년의 어떤 활동이나 어떤 그런 것들을 보면은 교육하고도 좀 연계교육이 영역이 크니까 그래서 그런 부분하고 연계해서 그렇게 이런 부분들이 만들어지면 청소년에 한정해서 하게 되면 조금 그 활용성이라든가 이런 것들이 좀 약하지

않을까 하는 생각이 들어서 연계해서 될 만들어 보시면 확장개념으로. 근데 이걸 좀 더 디자인을 한 번 좀 생각을 해 봐야 될 것 같아요.

7. 정부가 제가 알기로는 이제 데이터 3법, 개인정보 보호법, 정보화통신망법, 신용정보법을 사용해가지고 올 하반기에 이제 총괄컨트롤 타워를 출범한다고 들었습니다... 청소년 관련.. 관련해서는 우리 한국청소년정책연구원에 좀 지도하는 그런 노력이 필요하기도 하지 않겠는가

■ 기타

4. 지금 통계청이 또 모든 데이터를 그 자기네들이 받아서 관리를 하잖아요 대부분. 그 mds 가보시면 노동연구원 데이터도 들어와 있고 막다 들어와 있거든요. 장기적으로 다 관리한다는 거예요 통계청 얘기 들어봤더니. 그래서 그쪽 뭐 통계전문기관이기 때문에 저는 이걸 굳이 청소년

년 진흥원에서 청소년 데이터를 한다.. 하신다기 보다는 그냥 그 전문성이 훨씬 더 큰 통계청에서 하는 게 맞고요. 통계청이 해야 그 청소년 데이터를 평소에 안 쓰다가 갑자기 쓸 일이 생겨도 사실은 거기 가서 데이터를 알게 되고 그 데이터를 몰라서 못 쓰잖아요. 박사님 그럼 통계청 다른 자료 찾다가도 아 통계 청소년 자료도 있구나라는 걸 알게되고 활용이 되기 때문에 통계청이 뭐 홍보를 강화하는 측면이나 품질 통계 관리하는 측면이나 이런 것들이 봤을 때 저는 통계청에서 컨트롤타워로 역할을 하는 게 필요하다고 보고요.

6. 필요하다고 생각합니다. 여가부가 하고 싶어하겠죠? 안그러면 행안부죠. 공공데이터 전략위원회도 만들고 어쨌든 공공데이터 개방 관련된 과도 있고 하니까 행안부가 하는게 힘을 받을 수 있겠죠.

법으로 주계끔 명시가 돼있어야지

그거는 법으로 풀어야지 그거는 부처에서 요구를 다른 뭐 국무조정실이든 BH든 다른 쪽에서 요구를 한다고 되는 부분이 아닌 것 같아요

청소년이라는 키워드 들어가면 여가부예요 아동으로 만약에 청소년을 아동의 범주안에 넣으신다라고 하면 그건 복지부고.

8. 어떤 형태로든 필요하죠 당연히 있어야 된다고 생각합니다.

뭐 제가 잠깐 언급했다시피 뭐 그게 국무총리실이고 뭐 이런 식으로 가면 국무총리실 뭐 한도 끝도 없을 것이고 전부 다 그렇게 가야되는데 어.. 우선 1단계는 어떤 중추적인 기관에서 이제 센터의 형태로든 빅데이터 센터. 청소년 빅데이터 센터라는 형태로 출발을 하고 그다음에 이것이 이제 뭘니까 조금씩 조금씩 커져가지고 이제 그야말로 조금씩 발전되던가 해서 하나의 핵심기관이 따로 되든가...어느 기관이든 간에.

(4) 청소년 빅 데이터 체계 구축을 위한 중장기 계획 과정

청소년 빅 데이터 체계 구축에 있어 중장기 계획으로 필요한 과정을 묻는 질문에 대해 일차적으로 '중장기'의 시간개념을 정확히 할 필요성이 제기되었고, 또 다양한 의견 수렴의 필요성이 강조되었는데, 구체적으로 현장 교사들, 컴퓨터 사이언스 분야 등의 전문가들, 청소년정책 전문가들, 교육 및 정책 전문가들을 참여시킬 것을 조언하였다. 나아가 온톨로지의 지속적 보완이 요구된다고 하였고 마지막으로 절차의 중요성이 강조하였는데 구축대상 자료 확인, 기술적 정비, 자료 취합, 통계 품질관리 및 제공, 마지막으로 청소년 정책수요에 기초한 자료 발굴의 단계를 지적하였다. 이를 간단히 문제점 파악과 마스터플랜 또는 프레임구축으로 간략화한 응답도 있었고, 빅데이터 컨트롤타워를 시작으로 기반조성, 정보화 총괄, 빅데이터 팀 운영 강화 및 빅데이터 전문인력 확충, 빅데이터 활용으로 진행되어야 바람직하다는 견해도 있었다. 그리고 그 모든 과정에서 공감대 형성이 필요하고 효율적인 시범사업을 거쳐 중장기 계획으로 나아가야 한다는 조언이 있었다. 결국 구체적인 중장기 계획이 무엇인가 보다는 어떻게 계획을 초기 수립 단계부터 합리적이고 생산적으로 진행하는가가 중요하다는 견해가 지배적이었다.

<p>질문 19. 청소년 빅데이터 체계 구축을 위한 중장기 계획으로 어떠한 과정을 거치는 것이 바람직하다고 생각하시는지요?</p>	
<p>■ 개념 문제</p> <p>1. 중장기라고 보시면 몇 년 정도 생각하시는건지 조금 궁금하기도 하고 어떤 계획을 갖고 계신건지 또 2030 정도 생각하면 되는 걸까요?</p>	<p>■ 다양한 의견 수렴</p> <p>1. ... 그런 어떠한 그런 결정을 내리시거나 할 때 다양한 그런 곳에 계신 필드에 계신 선생님들의 의견도 종합할 필요가 있겠다는 생각이 조금 이제 들었어요.</p>

2. 그 전문가들 뭐 그런 컴퓨터 사 이언스라던가 이런 분들의 의견을 좀 더 많이 들어보셨으면 하는 생각이 네 들어요. ... 그 전문기관들의 보고서나 이런 거는 빅데이터만 가지고 얘기하는 거는 거의 없고요. 디지털 트랜스포메이션을 어떻게 할거냐 하는 부분이어서 그런 전반적인 디지털 전환에 대한 언급은 좀 해주실 필요가 있지 않겠나 싶은 생각이..

6. 빅데이터 전문가들 뿐만 아니라 청소년 그 연구자분들 박사님을 비롯한 청소년정책 전문가들 그리고 교육 쪽 전문가들에게 필요 특히 정책 쪽 하시는 그분들에게 쪽 받으셔가지고 정리를 해야지 않을까 싶어요. 이 중장기 계획은..

■ 지속적 보완

3. 매년 이 온톨로지에 대해서 업데이트를 해야된다고 생각해요 왜냐면 사회현상은 계속 바뀌니까. 그거

에 따른 요소도 계속 바뀔거라 생각 하거든요 ... 온톨로지도 또 몇년이 지나면 계속 업데이트가 되면서 구성이 바뀔 필요가 있다 그게 좀 증장기 적으론 중요한 것 같아요

■ 절차의 중요성

4. 또 원론적인데요. 일단은 구축 대상 자료를 확인하고 그 다음에 기술적인 정비해야 되고 그 다음에 그게 끝나면 이제 자료를 취합하고 그 다음에 통계 품질관리하고 제공하고 그 다음에 궁극적으로는 청소년 정책 수요에 기초한 자료를 발굴하고 그게 그냥 단계죠.

5. 실태 파악과 그니까 현황 문제점 이런 것들을 이제 파악하는 게 우선적으로 중요하고요. 그다음에 마스터플랜을 짜야 될 것 같습니다 ...그래서 일단은 전체를 알고 그 다음에 지금 뭐 일련 연차적으로 구축하는 거..

6. 가장 우선적으로 프레임구축 그

전체적인 이 보건 의료 빅데이터를 가지고 우리가 무엇을 할 것인지에 대한 그 큰 틀이 문제 18번에서 그 큰 틀을 작업을 하시는 연구인거 같고 올해에. 그래서 그 큰 틀이 잘 나왔으면 좋겠어요. 큰 틀 안에서 어떤 게 우선순위로 가야될지가 보일 것 같아요 예를 들어서 품질관리 얘는 나중에 중장기 계획인 거예요 데이터도 지금 모아지지도 않았는데 데이터가 뭐가 있는지도 모르는데 일단 데이터 수집이 우선적으로 해야 되는 나중에 중장기 계획은 결국엔 품질관리 쪽으로 가더라고요

7. 그니까 기반조성이 첫 번째 되면은 정보화 총괄하고 빅데이터 팀 운영 강화가 돼가지고 빅데이터 전문 인력 확충과 이런 걸 하려면 수요에 즉 빅데이터 활용 이런 것들이.. 그 다음에 이제 선행이 돼야 될 거고 더 나아가서는 빅데이터 전담부서가 좀 생겨야 된다고 생각을 하는데 그거는 빅데이터 컨트롤타워 역할을 수행하기 위해서...

8. 저는 어 이게 공감해서 그 뭘니까. 그게 제일 중요하잖아요. 공감이 얼마나 하는가 공감대형성 그니까 우선 그런 것을 우선 알려야 되겠죠. 교육부터 해야 되겠고 먼저 전문가 그룹에서 기본적인 중요성을 알리는 계획을 짜고 그걸 교육을 통해서 알리고 어느 정도 의견을 수렴한 후에 체계적인 계획을 수립하고 그니까는 아마 선제조건은 좀 이렇게 좀 공유되어야 해요

9. 일차적으로 지금 현재로서는 시범사업을 잘해보야 한다고 생각해요. 각각 저걸 만들어서 성공사례를 만들어내고 그런 것들을 여러 번 하면서 유관기관 간에 그걸 모아서 같이 하는 게 효율적이라는 걸 보여주고 그렇게 되면 사람들이 모여서 하다보면 점점 이쪽에서 주도권을 가져가는 그런 방식으로 해야 된다고 생각해요.

10. 지금까지 말씀드려서 뭐 특별히 드릴 말씀은 없는데요 뭐.. 어쨌든 고생하시는 것 같아요...

9) 청소년 빅데이터 체계 구축 및 활용의 발전방향

청소년 빅데이터 체계 구축 및 활용의 발전방향과 관련하여 응답자들은 질문을 빅데이터를 어떻게 발전적으로 활용될 수 있는가의 문제로 받아들이거나 어떻게 해야 더 발전할 수 있는가의 의미로 받아들이는 두 부류로 나누어졌다. 전자의 경우는 자료제공 자체 만이라도 발전적 기능이라는 입장, 빅데이터를 기존의 자료보다 더 강력한 증거자료로 활용할 수 있고 효과적이고 효율적인 정책지원을 할 수 있다는 입장이 있었다. 후자로는 디지털 전환 전략 속에서 빅데이터 체계 구축도 같이 볼 필요가 있다는 의견, 그리고 실질적 측면에서 데이터 관리 및 플랫폼 구축, 데이터의 품질관리 및 데이터 공유 플랫폼 구축이 필요하다는 입장이 있었다. 나아가 인식적 차원의 공감대가 요구된다는 입장도 있었다. 종사인력의 빅데이터에 관한 인식과 이해도를 높이고 연계와 연대와 이해의 폭을 넓힐 때 자살을 학교폭력에 대한 희생자 불링의 문제 등 다양한 사안에 대한 선제적 대응이 가능하다는 것이다. 마지막 방법론적으로는 종합보고서를 만들 필요가 있다는 입장, 그리고 청소년 정책 및 청소년 연구자 또 청소년 관련 정보의 수요자, 필요조사가 요구된다는 등의 입장이 있었다.

질문 20. 청소년 빅데이터 체계 구축 및 활용의 발전방향은 무엇이라고 생각하십니까?

■ 발전적 활용가능성

1. 증거자료로의 활용

1. 이게 청소년 관련해서 아까 박

사님께서 이걸 청소년 문제라고 생각하셨던 이유도 이제 그런 쪽에 어떤 해결방안을 제시할 때 증거자료로 활용하고자 하는 그런 것들이 있지 않았을까 ... 전통적인 연구방법

예를 들어서 질적 연구에서 몇사람을 포커스로 인터뷰를 하거나 아니면 양적연구에서 내가 관심이 있는 표본만 몇 사람을 뽑아서 일부 변인 가지고 하는 연구보다는 훨씬 더 파워풀 하게 어떤 증거가 될 수 있지 않을까 그런 생각에선 빅데이터 구축이 필요하다는 생각이..

2. 자료제공 및 정책지원

6. 결국은 데이터개방으로 가더라도 ...만약에 정부 차원에서 청소년 빅데이터라는 걸 딱 해서 데이터셋을 개방을 여러 개를 해준다라고 하면 그런 데이터셋 개방해 가지고 정말 멋진 어떤 사업이 나올 수도 있는거죠. 민간부분에서는 그리고 구분하셔야 될 것 같아요. 공공..물론 당연히 민간 부분에서의 개인정보는 이제 빅데이터라 해서 철저히 관리하고 뭐 그런 부분을 써주셔야 될 것 같고 공공차원에서는 정책지원이죠? 더 효과적인 더 효율적인 정책

지원을 하기 위해서 필요한..

■ 발전을 위한 방향

1. 디지털 전환

2. 정책의 그 증거기반을 마련한다든가 예측이라든가 서비스향상이라든가 뭐 이런 쪽이겠지만 여기에 빅데이터 하나만 분석해가지고 되는 일은 없거든요. 그럼 이제 나머지 그 기술적인 부분과 또 마인드셋이라든가 뭐 이런 것들을 그죠. 그 다음에 조직의 변화라든가 프로세스의 계산이라든가. 예. 이런 것들이 그 전부 다 디지털 전환과 관련되어 있는 건데 디지털 전환 전략 속에서 빅데이터 체계 구축도 같이 볼 필요가 있다...

7. 데이터 관리 및 플랫폼 구축이 필요하고 또 데이터의 품질관리도 필요하고 그리고 데이터 공유 플랫폼도 구축하는 것들이 필요하다고 생각이 됩니다.

2. 인식적 차원의 공감대

8. 활용의 발전방향 이거는 빅데이터가 청소년에서 얼마나 중요한가에 대해서 사람들이 좀 많이 알 필요가 있어요 그 대표적인 것 중 하나가 우리가 청소년 문제 이슈라든가 청소년 삶의 질에 관심이 많은데 빅데이터를 잘 활용하면 그 뭐 자살률 학교폭력에 대한 희생자 불링의 문제 이런 거에 대해서 우리가 미리 선제적인 이제 대응을 할 수 있단 말이에요 ... 그걸 우리가 정말 적절히 활용할 수 있도록 하는 연계와 연대와 이해의 폭을 넓히는... 이거 되게 중요하지 않는가 생각하고 아까 인력의 문제. 전문인력도 중요하지만 종사 인력의 빅데이터에 관한 인식에 박차, 이해도의 전가 중요합니다 그 다음에 더 중요한 거는 청소년 본인들. ...청소년 스스로가 자료를 관리하고 자료를 만들어내는 그 행위자라는 사실에 대한 자각이 있느냐 그런 게..

3. 선제적 대응을 위한 데이터 활용

3. 저는 사실 장기적으로는 청소년 사회에서 발생하는 문제에 대해서 가능하면 그 문제가 선제적으로 해결될 수 있는 게 제일 중요하다고 생각을 하는 사람이라서 어 그 이런 데이터분석을 한다는 건 사실 그런 목적이 더 동반돼야 된다고 생각을 해요. 그래서 가능하면 뭐 자살률이 떨어진다는지 아니면 뭐 학생들끼리 생기는 문제들 이런 것들 풀린다는지 뭐 애들에 대해 애들이 정말 필요로 하는 복지 서비스가 뭔지 라던지 뭐 그런 것들에 대해서 미리 구체적으로 제시할 수 있는 어 체계가 될 수 있는 게 중요하다 그렇게 생각합니다.

4. 방법론

9. 저는 그래도 일차적으로 제가 보기에는 그 할 수 있는 건 현실적으

로 뭘까 종합보고서를 만드는 거예요. 그거를 어떻게든 청소년쪽에 있는 사람들이 뜻을 모아서 저는 해야 된다고 생각해요. 그 하여튼 현재 할 수 있는 거는 지표를 모아서 할.. 모을 수 있는 거는 모으고 그 다음 나머지 안되는 것도 계속해서 수집하고, 그런데 문제는 이제 이거를 추진하는 주체가 불분명한게 그건 큰 문제예요.

10. 여기에 제일 중요한 건 뭐냐면 어.. 청소년 정책 또는 청소년 연구자 또 청소년 관련된 정보가 필요한 사람들의 니즈에 대한 것들이 한 챗터가 있으면 좋겠고 ... 그런 니즈를 담아낼 수 있는 그 일종의 편리하

고 효율적인 데이터를 어떻게 모으고 어떻게 활용한 건지에 대한 아이디어가 지금 이 파트가 그거거든요 그러면 데이터를 모을 때 데이터를 모으기 편리하도록 하기 위해서 현재 어디 각 기관에서 모은 데이터가 어떻게 있느냐를 분석하고 그중에 가져올 수 있는 데이터는 이거고 못 갖고 오는 것도 있을 거예요. 분석하다보면 아까 아카이브 뭐 이런 걸 하셨다니까. 그래서 그걸 분석해내고 그리고 어.. 그걸 기반으로 일차적으로는 이런 이런 서비스가 가능하고 또 이차적으로는 좀 더 이렇게 시간이 되면 이런 게 가능하고 이렇게 서비스 그렇죠.

3. 소결

청소년 빅데이터 체계구축을 위한 응답자들의 의견제시를 통해 생산적인 결과를 도출할 수 있었다. 먼저 빅데이터의 의미에 관련하여 응답자들은 빅데이터가 기존 데이터베이스의 처리방식을 초월하는 방대한 양의 정형 및 비정형 데이터라는 일반적 정의에 모두 공감하고 있었다. 또한 중요성과 관련해서는 모든 일상의 데이터화, 데이터의 확장과 연계 그리고 데이터 처리 기술의 발전 등으로 사회현상, 특히 청소년이나 청소년 활동과 관련된 사안의 보다 명시적인 확인, 정밀한 파악과 분석이 가능하게 되었고, 이를 통해 실증적 데이터를 기반으로 한 이론이나 논리의 발견과 예측이 가능해졌음을 지적하였다. 이러한 응답결과는 빅데이터, 무엇보다 청소년 관련 빅데이터 체계 구축의 필요성과 가능성 및 활용가능성을 내포하고 있다는 점에서 시사점이 있다. 다음 청소년 빅데이터의 개념과 관련하여 응답자들은 기존의 연구를 통해 알려진 다양성(Variety), 속도(Velocity), 휘발성(Volatility), 규모(Volume), 가변성(Variability), 정확성(Veracity), 그리고 가치(Value) 및 시각화(Visualization)의 8v가 모두 그 속성으로 유의미하다는 점은 인정하고 있었다. 다만 형식적 측면에서 8v를 모두 개념 안에 기술하기보다는 3v정도의 핵심속성을 정하고 나머지는 분리하여 보충설명으로 하면 좋겠다는 의견이 있었다. 내용적으로는 청소년 빅데이터의 개념이 청소년의 부정적 문제에만 국한되기보다 긍정적 문제를 포괄한 전체적인 방향으로 정리할 것을 건의하였다. 나아가 청소년 빅데이터의 본질을 ‘기술 및 능력’으로보다는 ‘정보나 자원’ 정도로 규정하는 것이 더 합리적일 것이라는 의견이 있었다. 이상 응답자들의 견해 가운데 합리적이며 타당한 부분을 고려하여 본 연구는 청소년 빅데이터의 정의를 다음과 같이 수정 정리하였다.

청소년 빅데이터는 ‘청소년 삶의 제반 영역에서 다양한 형태(Variety)로 빠르게 (Velocity) 생성되는, 대규모(Volume) 데이터 가운데, 청소년이 현실적으로 직면

하는 다양한 사안의 적시성 있고 적절한 방향제시를 위해, 유의미한 데이터로 선별, 취합 및 선택적으로 분석되는 모든 데이터'이다.

다음으로 국내·외 빅데이터 체계 구축 상황과 우수사례에 대해 응답자들은 국내의 경우 교육부의 에듀 데이터 서비스, 서울시의 버스 노선 및 네비게이션 등 주로 공공 데이터를 언급하였고, 국외의 경우는 캐나다의 자살위험 예측 프로젝트 등이 제시되었다. 그리고 빅데이터 국내 우수사례로는 복지 사각지대 발굴 시스템, 위기 아동발굴 시스템 그리고 보건의로 빅데이터가 거론되었으며 국외 우수사례로는 페이스북과 유튜브를 분석해주는 미국 사이트 빅팟나인이 지목되었다. 나아가 청소년 빅데이터 체계 구축 및 활용의 구체적인 우수 사례와 관련해서는 국내의 경우 2013년 여성가족부의 위기 청소년 빅데이터, 교육부의 학생생활기록부 사이트 등이 언급되었고 국외의 경우는 유스 프로그램 데이터 베이스인 best practice가 거론되었다. 또 빅데이터의 활용 경험은 대체적으로 개인적인 연구 활동이나 정부부처 등의 의뢰가 다수였다. 연구로는 머신러닝기법들을 적용하여 다양한 패널의 기존 데이터를 진료와 의사결정나무 등의 주제에 적용한 경우 등이었고 정부부처 의뢰사업으로는 빅데이터 플랫폼인 센터 구축사업 등이 있었다. 그 외 응급의료서비스, 보안, 정보시스템 등 다양한 사회문제 해결을 위한 사업에 참여한 것을 확인할 수 있었다. 현재 활용 가능한 정형 빅데이터에 대해 응답자들은 정보화진흥원의 공공 빅데이터들, 통계청 데이터 등 연구기관의 연구 자료들과 정부부처의 각종 데이터들을 언급하였다. 비정형 빅데이터로는 SNS 소셜 데이터, 유튜브 등 동영상 데이터, 사진데이터 등이 거론되었고, 특별히 청소년 데이터로는 여가부의 청소년 종합 실태 등이 언급되었다.

청소년 빅데이터가 실질적으로 다루어야 할 사안에 대하여 응답자들은 크게 생애설계적 관점과 생애수정적 관점의 두 관점에서 접근하는 모습을 보여주었다. 생애설계적 관점, 즉 청소년의 인생을 생산적이고 긍정적으로 구성하는 측면에

중심을 둔 입장은 중분류로 진로, 주관적 복지감, 심리, 안전, 인구, 가족, 건강, 교육, 여가, 문화, 시간 이용, 공감역량, 관계 및 사회참여 등등을 언급하였다. 그리고 그에 대응하는 세부항목들로는 홀로 보내는 시간, 독서, 스마트폰 사용량과 사용시간, 규칙적인 운동 여부, 아침 식사 여부, 수면 시간, 비만도, 자원봉사활동 등이 있었으며, 가족 구성 상태에 관련한 것들로 한부모 또는 조손가정 여부, 가구소득 등이 언급되었다. 생애수정적 관점, 즉 청소년에게 부정적 영향을 미치는 것들과 관련해서는 스마트폰 중독이나 인터넷 중독, 분노 조절장애, 공격, 스트레스, 학교폭력, 자살, 불안, 비행, 자아존중감, 자아효능감, 자아탄력성 등이 거론되었고 세부항목으로 비대면 활동에 참여하는 시간, 학교 학원 이외의 하루 활동 디지털 활용시간 등이 언급되었다. 또 청소년 빅데이터에 꼭 들어가야 할 세부지표에 대해서도 응답자들은 생애설계와 생애수정의 두 관점에서 접근하였다. 전자는 진로, 직업 등과 관련하여 청소년들의 일과시간에 대한 관심이 있었고, 특히 긍정적 지표 개발의 중요성을 강조하기도 하였다. 생애수정적 관점의 지표로 생각하는 키워드로는 비행행동, 가출, 자살, 학교 밖 청소년, 미디어중독, 청소년 정서 행동장애, 가정폭력, 부모자녀갈등, 소비문제, 신체적 정신적 건강, 무엇보다 비교의식 등 청소년 관련 대부분의 문제가 거론되었다.

다음으로 청소년복지 증진 및 정책개발, 미래예측을 위해 청소년 빅데이터를 통해 우선적으로 분석, 도출되어야 하는 인과관계(독립변인, 종속변인)에 대하여 독립변인으로 미래나 삶 등에 대한 가치관을, 그 외 일반적인 독립변인으로 많은 키워드를 제시해주었다. 지역, 성별, 연령, 학년, 학교 사교육, 가정 배경, 가족과 관련된 부모 사별, 새부모 재혼여부 등이 있었고 종속변인은 부정적 경험 관련 키워드가 많았는데 구체적으로 임신, 낙태 중절, 성추행, 성폭력, 성관계, 우울증, 애정 결핍, 수면제, 정신치료, 자살상담, 항우울제, 경제적 어려움, 실직, 기초생활수급, 이성 관계 어려움, 인터넷중독 장애인, 질병, 외모, 분노, 스트레스, 불면증, 충동행동, 가출, 폭력, 퇴학, 자퇴 자해, 음주, 흡연, 약물, 본드사용, 위기 행동,

범죄 행동 등이 거론되었다.

다음 방법론적인 측면에서 청소년 빅데이터의 활용도를 높이기 위한 방안으로 응답자들 중 다수는 전처리의 중요성과 데이터 신뢰의 문제 그리고 데이터 개방을 지적했다. 먼저 비정형 데이터와 정형데이터 둘 다 데이터 전처리를 위해 시간과 비용이 많이 들어 효율성이 떨어지는 문제를 지적했다.

따라서 데이터 클리닝을 비롯 연구자 편의를 고려한 기초 가공이 이루어진다면 빅데이터 활용도가 상승할 것으로 진단했다. 다음으로 데이터 신뢰의 문제를 언급하였는데 특히 비정형데이터의 경우 일반화의 어려움과 타당성 있는 설명력의 결핍 가능성이 언급되었다. 따라서 신뢰성과 타당성 있는 빅데이터에 대한 주기적인 발표의 필요성을 제기하였다. 그리고 마지막으로 데이터 개방과 관련하여 필요한 자료에 대한 홍보 및 접근성이 높아져야 한다고 지적하였다. 데이터의 실질적인 적합성과 실효성을 높이기 위해서는 수용자가 필요로 하는 좋은 품질의 자료를 언제든지 검색할 수 있어야 한다는 것이다. 그 외 활용도를 높이기 위한 방안으로 풍부한 데이터의 양과 사용자의 고양된 데이터 마인드셋이 필요하다는 지적이 있었다. 나아가 다양한 청소년관련 빅데이터를 이용자 중심으로 체계화하기 위하여 기본적으로 온톨로지의 필요성, 이용자 중심의 체계화, 그리고 공유 시스템의 필요성이 강조되었다. 무엇보다 응답자들은 분산되어 있는 데이터들의 상호 개방과 연계를 통한 효율성 제고를 위하여 컨트롤 타워의 필요성에 공감하고 있었다. 다만 컨트롤 타워를 어디에 둘 것인가 하는 문제에 대해 청소년정책연구원이 말아야 한다는 다수 의견과 통계청에서 컨트롤타워 역할을 하는 것이 데이터 활용상 홍보 강화 그리고 품질 관리 측면에서 효율성을 제고할 수 있다는 입장이 있었다.

청소년 빅데이터 체계 구축 및 활용의 장점에 대해 응답자의 다수는 데이터 활용시의 효율성제고를 언급하였다. 특히 데이터 전처리에 필요한 비용과 시간절감 그리고 각 기관마다 보유하고 있는 자료의 특성들을 고려한 데이터 연계가능성의 장점을 지적하며 이를 통해 청소년 문제에 대한 선제적 대응, 예방과 청소년의

잠재력 개발 또 청소년의 편의개발 등 다양한 측면에서의 정책적 활용이 가능하리라고 보았다. 또 청소년 문제에 대한 정확한 진단을 통해 사안 관련 현황 파악과 증거기반 정책수립, 나아가 그러한 일련의 정책이 가져올 효과에 대한 예측 가능성이 가능하리라고 보았다. 그리고 이 모든 장점을 통해 결국 안전하고 건강한 청소년을 길러내기 위한 실질적 개선이 이루어질 수 있다고 진단했다. 반면 단점으로는 일차적으로 개인정보 유출을 통한 사생활침해 가능성이 우려되는 바, 개인정보 보호를 위한 확실한 법적 장치의 필요성이 지적되었다. 역으로 개인정보보호의 문제로 인한 정보접근의 어려움이 언급되기도 하였다. 또 데이터 품질관리나 신뢰성의 문제 및 막대한 비용문제 등이 제기되었다.

한국의 상황에서 청소년 빅데이터 체계 구축 및 활용의 한계점에 대한 질문에 대해 응답자의 다수는 데이터 확보 문제를 지적했으며 그 이유로는 1) 개인정보 보호 정책으로 인해 자유로운 정보활용의 한계, 2) 데이터공개 시점의 문제점, 데이터 확보를 어렵게 하는 3) 수집구조의 문제, 그리고 4)데이터 오용문제 등을 지적했다. 이는 결국 빅데이터의 구축과 관리의 중심이 되는 컨트롤타워의 필요성이 강조된 것이라고 볼 수 있다. 컨트롤타워가 있을 때 데이터 확보, 공유, 품질관리 및 서비스의 효율적 수행이 가능하다고 보는 것이다. 동시에 빅데이터의 생산적 활용을 위해 데이터 개방, 역량 있는 분석가의 양산, 담당공무원의 추진력, 그리고 수직구조 개선 등이 지적되었다.

청소년정책에 빅데이터를 어떻게 활용할 수 있을지에 대한 응답은 크게 두 가지였다. 첫째는 증거 기반 정책 수립이고 둘째는 실효성 있는 정책 개발이다. 전자와 관련하여 빅데이터는 실증자료를 활용한 정책 결정은 물론 정책에 대한 사후 평가를 가능하게 해줄 것으로 평가했다. 또 위기청소년이나 청소년의 정신건강 등 구체적인 사안들에 대하여 실효성 있는 정책, 즉 정책을 위한 정책이 아닌 실무적 시사점을 도출할 수 있는 정책개발에 기여할 것으로 평가했다. 그 외 기존의 연구방식이 주지 못하는 새로운 정보를 도출할 수 있고 청소년과 청소년 문제에

대한 관심을 유발할 수도 있으리라고 내다봤다. 이러한 청소년 대상 빅데이터 체계 구축을 위해 선결되어야 할 사항으로 응답자들의 다수는 자료의 통일적 체계 구축, 즉 온톨로지의 필요성을 지적하였다. 나아가 빅데이터 분석과 운영관리 체계 수립을 통한 빅데이터 분석체계와 빅데이터 통합관리 방안 수립, 그리고 빅데이터 인프라 환경분석과 도입계획을 수립하기 위한 시스템 환경 등이 필요하다고 보았다. 또한 가버넌스의 필요성, 구성원들의 디지털 친화적 내지는 데이터 친화적 마인드셋으로의 패러다임 전환, 그리고 필요성 및 활용성에 대한 가치 인식 공유, 빅데이터 구축과 데이터 수집에 있어 청소년과 학부모에 대한 인식 고려가 필요하다는 주장이 있었다. 나아가 빅데이터 체계구축에 있어 우선적으로 고려해야 할 주제로 응답자 다수는 청소년의 삶과 행복이라고 답하였다. 구체적으로는 청소년들의 시간활용, 가정 내 아동학대 및 폭력, 그리고 중독이나 자기조절 문제 등의 소극적 영역과 청소년의 신체적, 정서적 건강 및 안전 분야 인지적 영역의 학업성취도 등 적극적 영역의 주제들이 언급되었는데 특히 자율성의 주제가 강조되었다. 기타 실천적 주제로 봉사활동이나 사회적 기업 등 시민적 관여의 문제도 거론되었다.

또한 청소년 빅데이터 체계 구축에 있어 중장기 계획 수립을 위해 현장 교사들, 컴퓨터 사이언스 분야 등의 전문가들, 청소년정책 전문가들, 교육 및 정책 전문가들이 참여하는 다양한 의견 수립의 필요성이 강조되었다. 동시에 온톨로지의 지속적인 보완과 함께 절차의 중요성이 강조되었다. 즉 구축대상 자료 확인, 기술적 정비, 자료 취합, 통계 품질관리 및 제공, 마지막으로 청소년 정책수요에 기초한 자료 발굴의 단계를 지적하였다. 그리고 그 모든 과정에서 공감대 형성이 필요하고 효율적인 시범사업을 거쳐 중장기 계획으로 나아가야 한다고 보았다. 마지막으로 청소년 빅데이터 체계 구축 및 활용의 발전방향에 대하여 응답자들은 질문을 빅데이터의 발전적 활용가능성과 빅데이터의 발전 가능성이라는 두 가지 의미로 받아들였다. 전자의 경우는 자료제공 자체만이라도 발전적 기능이라는 입장, 빅데이터

를 기존의 자료보다 더 강력한 증거자료로 활용할 수 있고 효과적이고 효율적인 정책지원을 할 수 있다는 입장이 있었다. 후자로는 디지털 전환 전략 속에서 빅데이터 체계 구축도 같이 볼 필요가 있다는 의견, 그리고 실질적 측면에서 데이터 관리 및 플랫폼 구축, 데이터의 품질관리 및 데이터 공유 플랫폼 구축이 필요하다고 보았다. 나아가 인식적 차원의 공감대, 종사인력의 빅데이터에 관한 인식과 이해도 제고 등의 필요성이 지적되었다.

○———— 제6장 전문가 조사

- 1. 조사대상 일반특성
- 2. 빅데이터 사용 경험
- 3. 빅데이터 관련 욕구

1. 조사대상 일반특성⁷⁹⁾

전문가 조사는 기관생명윤리위원회(IRB)의 심의면제 확인을 득한 뒤 실시되었다(승인번호 202008-HR-고유-017).

조사대상 일반특성을 살펴보면 다음과 같다.

표 VI-1 전문가 조사대상 일반특성

구분		빈도	%	구분		빈도	%
성별	남자	31	83.8	직업	현장전문가	7	18.9
	여자	6	16.2		공무원	2	5.4
	계	37	100.0		교수, 연구원	28	75.7
			계		37	100.0	
학력	대학졸업	1	2.7	경력	10년미만	5	13.5
	석사재학/졸업	4	10.8		10년이상 20년미만	9	24.3
	박사재학/졸업	30	81.1		20년이상 30년미만	13	35.1
	기타	2	5.4		30년이상 40년미만	8	21.6
	계	37	100.0		40년이상	2	5.4
연령	20대	1	2.7		계	37	100.0
	40대	10	27.0		(평균/SD)	(20.3/10.5)	
	50대	18	48.6		(최소/최대)	(3/40)	
	60대	8	21.6				
	계	37	100.0				
	(평균/SD)	(53.4/8.2)					
	(최소/최대)	(29/67)					

79) '제6장 전문가 조사'는 한국청소년정책연구원 서정아 선임연구위원이 작성하였다.

전문가 조사 참여자는 37명으로 성별로는 남자가 83.8%, 여자가 16.2%, 직업별로는 현장전문가 18.9%, 공무원 5.4%, 교수, 연구원 75.7%였다. 학력은 대학졸업이 2.7%, 석사재학/졸업 10.8%, 박사재학/졸업 81.1%, 기타 5.4%였으며 연령은 20대 2.7%, 40대 27.0%, 50대 48.6%, 60대 21.6%이며 평균 연령은 53.4세였다. 경력을 살펴보면 10년 미만이 13.5%, 10년이상 20년 미만이 24.3%, 20년 이상 30년 미만이 35.1%, 30년 이상 40년 미만이 21.6%, 40년 이상이 5.4%로 평균 경력은 20.3년이였다.

2. 빅데이터 사용 경험

1) 공공 영역 빅데이터 사용 연구·분석 경험

① 공공 영역 빅데이터 사용 경험 유무

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 응답하였다. 전문가별로는 현장전문가, 공무원은 66.7%, 교수, 연구원은 46.4%가 사용 경험이 있다고 응답하였다.

표 VI-2 공공 영역 데이터 사용 경험 유무

구분		공공 영역 데이터 사용 경험 유무				계		x2
		있음		없음				
		빈도	%	빈도	%	빈도	%	
전문가 구분	현장전문가, 공무원	6	66.7	3	33.3	9	100.0	1.1
	교수, 연구원	13	46.4	15	53.6	28	100.0	
계		19	51.4	18	48.6	37	100.0	

*p<.05, ** p<.01, *** p<.001

② 사용한 공공 영역 빅데이터(기관 기준)

공공 영역의 빅데이터를 사용하여 연구 또는 분석을 실시한 경험이 있는 경우 사용한 공공 영역 빅데이터를 기관 기준으로 살펴보면 다음과 같다.

표 VI-3 사용한 공공 영역 빅데이터(기관 기준)(다중응답)

구분			전문가				계		
			현장전문가, 공무원		교수, 연구원				
			빈도	%	빈도	%	빈도	%	
사용한 공공영역 빅데이터 (기관 기준)	중앙 행정기관	통계청	2	40.0	6	50.0	8	47.1	
		경찰청	0	0.0	1	8.3	1	5.9	
		과학기술정보통신부	0	0.0	2	16.7	2	11.8	
		미래창조과학부	0	0.0	1	8.3	1	5.9	
		국도교통부	0	0.0	3	25.0	3	17.6	
		여성가족부	3	60.0	2	16.7	5	29.4	
		보건복지부	1	20.0	0	0.0	1	5.9	
		질병관리청	0	0.0	1	8.3	1	5.9	
		소방청	0	0.0	1	8.3	1	5.9	
		교육부 (시도교육청포함)	1	20.0	0	0.0	1	5.9	
	지방 자치단체	서울특별시	3	60.0	2	16.7	5	29.4	
		충청남도	0	0.0	1	8.3	1	5.9	
	공사/공단/ 공공기관	국민건강보험공단	0	0.0	1	8.3	1	5.9	
		한국콘텐츠진흥원	1	20.0	0	0.0	1	5.9	
		중소기업진흥공단	0	0.0	1	8.3	1	5.9	
		국립청소년수련원	1	20.0	0	0.0	1	5.9	
		서울문화재단	1	20.0	0	0.0	1	5.9	
	연구기관	한국청소년정책연구원	2	40.0	8	66.7	10	58.8	
		한국보건사회연구원	0	0.0	4	33.3	4	23.5	
		한국직업능력개발원	0	0.0	1	8.3	1	5.9	
		과학기술정책연구원	0	0.0	1	8.3	1	5.9	
	국제기구	국제연합(UN)	1	20.0	0	0.0	1	5.9	
	외국 행정/공공기관	미국공공기관	0	0.0	1	8.3	1	5.9	
		미국주정부	0	0.0	1	8.3	1	5.9	
	계			16	320.0	38	316.4	54	317.9

전체적으로 한국청소년정책연구원이 58.8%으로 가장 많았으며 다음으로 통계청 47.1%, 여성가족부 29.4%, 서울특별시 29.4%, 한국보건사회연구원 23.5%, 국토교통부 17.6%, 과학기술정보통신부 11.8%, 미래창조과학부, 질병관리청, 경찰청, 보건복지부, 충청남도, 소방청, 한국콘텐츠진흥원, 중소기업진흥공단, 국립청소년수련원, 서울문화재단, 교육부(시도교육청포함), 국민건강보험공단, 한국직업능력개발원, 과학기술정책연구원, 국제연합(UN), 미국공공기관, 미국주정부 각각 5.9% 순이었다. 현장전문가, 공무원의 경우 여성가족부와 서울특별시가 60.0%로 가장 많았으며 다음으로 통계청, 한국청소년정책연구원이 40.0%, 보건복지부, 교육부(시도교육청 포함), 한국콘텐츠진흥원, 국립청소년수련원, 서울문화재단, 국제연합(UN)이 각 20.0%였다. 교수, 연구원의 경우 한국청소년정책연구원이 66.7%로 가장 많았으며 다음으로 통계청 50.0%, 한국보건사회연구원 33.3%, 국토교통부 25.0%, 서울특별시, 여성가족부, 과학기술정보통신부 각 16.7%, 미래창조과학부, 소방청, 경찰청, 국민건강보험공단, 중소기업진흥공단, 질병관리청, 충청남도, 한국직업능력개발원, 과학기술정책연구원, 미국 공공기관, 미국 주정부가 각각 8.3%였다.

전문가가 사용한 공공 빅데이터의 중요 키워드를 살펴보면 다음과 같다.

112신고, 스마트폰과의존, 신규일자리 도출, 건강보험코호트, 게임, 정신건강, 문화지표, 데이터기반, 과학행정, 청년, 사회경제실태, 자원봉사, 청소년종합실태, 청소년인권실태, 민관융합데이터, 다문화가족실태, 지역아동센터, 지역사회건강, 청소년 성매매, 아동종합실태조사, 수련활동참여, 동아리운영, 아동청소년패널, 다문화청소년패널, 인터넷중독, 학생통계, 학교통계, 인구주택총조사, 청소년상담사, 주거실태, 청년패널, 의료패널, 복지패널, 주택실거래가, 사회조사, 지역별 고용조사, 응급의료출동(119)

③ 공공 영역 빅데이터로 분석한 내용/주제 영역

공공 영역 빅데이터로 분석한 내용/주제 영역을 살펴보면 다음과 같다.

표 VI-4 공공 영역 빅데이터로 분석한 내용/주제(다중응답)

구분			전문가				계		
			현장전문가, 공무원		교수, 연구원				
			빈도	%	빈도	%	빈도	%	
공공 영역 빅데이터로 분석한 내용/ 주제 영역	인구 및 가족	아동·청소년 인구 규모	0	0.0	1	9.1	1	6.3	
		부모자녀관계	0	0.0	4	36.4	4	25.0	
		기타	0	0.0	2	18.2	2	12.5	
	교육 및 역량	기타	0	0.0	1	9.1	1	6.3	
	건강 및 심리	신체건강	1	20.0	2	18.2	3	18.8	
		영양섭취	0	0.0	1	9.1	1	6.3	
		정신건강	0	0.0	2	18.2	2	12.5	
		중독	1	20.0	2	18.2	3	18.8	
		기타	0	0.0	3	27.3	3	18.8	
	활동 및 문화	동아리활동	1	20.0	0	0.0	1	6.3	
		자원봉사활동	0	0.0	1	9.1	1	6.3	
		여가생활	1	20.0	0	0.0	1	6.3	
		기타	2	40.0	4	36.4	6	37.5	
	복지 및 권리	취약, 위기청소년	0	0.0	1	9.1	1	6.3	
		복지시설 이용	1	20.0	0	0.0	1	6.3	
		복지서비스 이용	1	20.0	1	9.1	2	12.5	
		권리 및 참여 의식	1	20.0	0	0.0	1	6.3	
		기타	0	0.0	1	9.1	1	6.3	
	비행 및 범죄	음주/흡연/약물	0	0.0	1	9.1	1	6.3	
		성관련 경험	1	20.0	0	0.0	1	6.3	
	기타	아동청소년 이외 영역	1	20.0	10	90.9	11	68.8	
	계			11	220	37	336.6	48	300.8

전체적으로 아동청소년 이외 영역이 68.8%로 가장 많았으며 다음으로 활동 및 문화 기타 37.5%, 부모자녀관계 25.0%, 중독, 신체건강, 건강 및 심리 기타 각각 18.8%, 정신건강, 복지서비스 이용 각각 12.5%, 동아리활동, 영양섭취, 여가생

활, 복지 및 권리 기타, 취약, 위기청소년, 복지시설 이용, 아동청소년 인구 규모, 권리 및 참여 의식, 음주/흡연/약물, 성관련 경험, 자원봉사활동이 각각 6.3%였다. 전문가별로는 현장전문가, 공무원은 활동 및 문화 기타가 40.0%로 가장 많았으며 다음으로 신체건강, 중독, 동아리활동, 여가생활, 복지시설 이용, 복지서비스 이용, 권리 및 참여 의식, 성관련 경험, 아동청소년 이외 영역 20.0% 순이었다. 교수, 연구원은 아동청소년 이외 영역이 90.9%로 가장 많았으며 다음으로 활동 및 문화 기타 36.4%, 부모자녀관계 36.4%, 건강 및 심리 기타 27.3%, 인구 및 가족 기타, 신체건강, 정신건강 각 18.2%, 교육 및 역량 기타, 영양섭취, 아동청소년 인구 규모, 자원봉사활동, 음주/흡연/약물, 취약 위기 청소년, 복지 및 권리 기타, 복지서비스 이용이 각각 9.1%였다.

공공 영역 빅데이터로 분석한 내용/주제의 중요 키워드는 다음과 같다.

가정폭력, 성폭력, 폭력, 핫스팟, 지역특성, 게임사용, 청소년건강, 예술교육, 수요자, 지역 문화환경, 워크넷, 구직자, 구인업체, 직업군 분류 및 개발, 직무/직업도출, 유동인구, 공공 와이파이, 정부간기구, 청(소)년권리와 책임규정, 청소년기본권, 정신건강관련 서비스이용 실태, 만족도, 지역, 청소년시설, 행복, 청년취업현황, 그릿(Grit), 청소년 사회참여, 이직, 직업의식, 청소년실태, 교육만족도, 자원봉사, 발전전략, 활성화, 대사중후군, 변화계적, 대졸 청년, 직무만족도, 가정폭력신고사건, 발생분석, 날씨, 기온, 습도, 라이프스타일 분석, 스트레스, 다문화가족, 정신건강관련 서비스, 학업성취, 비만, 아동청소년돌봄, 청소년 성매매 실태, 사회관계망, 소비활동, 자조집단, 중학생, 진로결정, 청소년동아리, 청소년활동 참여, 청소년비행, 다문화가족 차별경험, 주택가격, 청소년지도인력, 청소년상담사, 다문화가족, 자아존중감, 의료패널, 의료과다이용, 예측모형개발, 인터넷중독위험 예측, 고용 창출, 연구개발활동, 응급의료서비스, 수요와 공급, 생산성 증대, 경제효과, 금융위기

④ 공공 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항

공공 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항을 살펴보면 다음과 같다. 전체적으로 개인정보 보호로 인한 접근의 제한성이 37.5%로 가장 많았으며 다음으로 데이터 형식간의 통일성 부재로 인한 분석상의 어려움, 데이터 품질 및 개선사항에 대한 의견수렴 부족, 홍보 부재로 인한 데이터 관련 정보 부재,

데이터 세부 사항, 범례 등에 대한 설명 부족, 원스톱서비스 또는 센터의 부재(한 곳에 모아놓기를 희망), 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, 서비스 간 연계 프로세스 부재가 각 6.3%이었다.

표 VI-5 공공 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항(다중응답)

구분	전문가				계		
	현장전문가, 공무원		교수, 연구원				
	빈도	%	빈도	%	빈도	%	
공공 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항 (다중응답)	개인정보 보호로 인한 접근의 제한성	0	0.0	6	54.5	6	37.5
	데이터 형식 간의 통일성 부재로 인한 분석 상의 어려움	1	20.0	1	9.1	2	12.5
	데이터 간의 연계 가능성 희박	0	0.0	1	9.1	1	6.3
	홍보 부재로 인한 데이터 관련정보 부재	0	0.0	2	18.2	2	12.5
	데이터 항목의 누락이나 중첩 등 부정확성 존재	0	0.0	1	9.1	1	6.3
	자료의 복잡성	1	20.0	0	0.0	1	6.3
	원데이터를 분석 가능한 데이터로 변환하는 과정 불편	0	0.0	2	18.2	2	12.5
	데이터 표집의 한계성 (예: 편의추출 등)	0	0.0	1	9.1	1	6.3
	데이터 세부사항, 범례 등에 대한 설명 부족	1	20.0	1	9.1	2	12.5
	필요한 데이터의 부재(예: 청년)	1	20.0	0	0.0	1	6.3
	원스톱서비스 또는 센터의 부재 (한 곳에 모아놓기를 희망)	1	20.0	1	9.1	2	12.5
	데이터품질 및 개선사항에 대한 의견 수렴 부족	0	0.0	2	18.2	2	12.5
	신뢰도 및 타당도에 대한 확인의 어려움	0	0.0	1	9.1	1	6.3
	분석과 정책, 서비스간 연계 프로세스 부재	1	20.0	0	0.0	1	6.3
계	6	120	19	172.8	25	156.6	

전문가별로는 현장전문가, 공무원은 자료의 복잡성, 데이터 형식간의 통일성 부재로 인한 분석상의 어려움, 분석과 정책, 서비스간 연계 프로세스 부재, 필요한 데이터의 부재(예: 청년), 데이터 세부사항, 범례 등에 대한 설명 부족, 원스톱 서비스 또는 센터의 부재(한 곳에 모아놓기를 희망)가 각 20.0%였다. 교수, 연구원은 개인정보보호로 인한 접근의 제한성이 54.5%로 가장 높았으며 다음으로 원데이터를 분석가능한 데이터로 변환하는 과정 불편, 데이터 품질 및 개선 사항에 대한 의견수렴 부족, 홍보부재로 인한 데이터 관련 정보부재가 각 18.2%, 데이터 형식간의 통일성 부재로 인한 분석 상의 어려움, 데이터 간의 연계 가능성 희박, 데이터 항목의 누락이나 중복 등 부정확성 존재, 데이터 표집의 한계성(예: 편의추출 등), 데이터 세부사항, 범례 등에 대한 설명 부족, 신뢰도 및 타당도에 대한 확인의 어려움, 원스톱 서비스 또는 센터의 부재(한곳에 모아놓기를 희망)가 9.1%였다.

공공 영역의 빅데이터를 사용하면서 불편하거나 개선했으면 하는 점 중 중요한 것을 살펴보면 다음과 같다.

- 데이터 수집시 입력 항목을 현실에 맞도록 보다 정교하게 구분하고 항목간 배타성과 포괄성을 유지하도록 개선되어야 함.
- 청소년시설 평가 등에 대한 양적, 질적 조사 현황에 대한 데이터를 종합적으로 알 수 있는 청소년종합관리서비스망이 한국청소년정책연구원, 한국청소년활동진흥원, 한국청소년상담복지개발원, 통계청 등에 구축되어 있지만 청소년활동, 청소년복지, 청소년보호(성보호 포함)을 종합적으로 빅데이터 관리할 수 있는 1개의 아동청소년빅데이터 센터(가칭)를 구축하는 것이 필요함.

2) 민간 영역 빅데이터 사용 연구·분석 경험

① 민간 영역 빅데이터 사용 경험 유무

민간 영역 빅데이터 사용 경험을 살펴보기 위해 “포털이나 SNS, 인터넷 등 민간 영역의 빅데이터를 사용하여 연구 또는 분석한 경험”을 질문한 결과 전체적으로 현장전문가, 공무원 9명, 교수, 연구원 26명이 경험이 있다고 응답하였다. 현장전문가, 공무원의 경우 44.4%, 교수, 연구원의 경우 25.9%가 민간 영역 빅데이터를 사용하여 연구 또는 분석을 실시한 경험이 있는 것으로 나타났다.

표 VI-6 민간 영역 빅데이터 사용 경험 유무

구분		민간 영역 데이터 사용 경험 유무				계		x2
		있음		없음				
		빈도	%	빈도	%	빈도	%	
전문가 구분	현장전문가, 공무원	4	44.4	5	55.6	9	100.0	.95
	교수, 연구원	7	26.9	19	73.1	26	100.0	
계		11	31.4	24	68.6	35	100.0	

*p<.05, ** p<.01, *** p<.001

② 사용한 민간 영역 빅데이터

민간 영역 빅데이터를 사용한 경험이 있는 경우 사용한 민간 영역 빅데이터의 유형을 살펴보면 전체적으로 국내일간지/온라인뉴스채널, SK텔레콤+통계청 유동생활인구분석, 복합데이터(뉴스+게시판+자료실+SNS), 다음/네이버카페가 27.3%로 가장 많았으며 다음으로 금융빅데이터플랫폼(BC 카드) 18.2%, 네이버지식인, 부동산114, 네이버랩, 트위터, 유튜브, 다음SNS, 빅카인즈, 구글트렌드, 인스타그램, 구글크롬로그파일(log file), 네이버블로그, 페이스북, 각종온라인게시판/댓글, 네이버판이 각 9.1%로 나타났다.

표 VI-7 사용한 민간 영역 빅데이터 유형(다중응답)

구분	전문가				계		
	현장전문가, 공무원		교수, 연구원				
	빈도	%	빈도	%	빈도	%	
사용한 민간 영역 빅데이터 유형 (다중응답)	구글트렌드	0	0.0	1	14.3	1	9.1
	금융빅데이터플랫폼(BC 카드)	2	50.0	0	0.0	2	18.2
	네이버랩	1	25.0	0	0.0	1	9.1
	네이버지식인	1	25.0	0	0.0	1	9.1
	다음/네이버카페	0	0.0	3	42.9	3	27.3
	부동산114	0	0.0	1	14.3	1	9.1
	국내일간지/온라인뉴스채널	0	0.0	3	42.9	3	27.3
	유튜브	0	0.0	1	14.3	1	9.1
	트위터	0	0.0	1	14.3	1	9.1
	네이트판	1	25.0	0	0.0	1	9.1
	다음SNS	0	0.0	1	14.3	1	9.1
	빅카인즈	1	25.0	0	0.0	1	9.1
	SK텔레콤+통계청 유동생활인구분석	3	75.0	0	0.0	3	27.3
	인스타그램	0	0.0	1	14.3	1	9.1
	구글크롬로그파일(log file)	0	0.0	1	14.3	1	9.1
	네이버블로그	0	0.0	1	14.3	1	9.1
	페이스북	0	0.0	1	14.3	1	9.1
	각종 온라인게시판/댓글	0	0.0	1	14.3	1	9.1
	복합데이터 (뉴스+게시판+자료실+SNS)	1.00	25.0	2.00	28.6	3	27.3
	계	10	250	18	257.4	28	254.8

전문가별로는 현장전문가, 공무원은 SK텔레콤+통계청 유동생활인구분석이 75.0%로 가장 많았으며 다음으로 금융빅데이터플랫폼(BC 카드) 50.0%, 네이트판, 네이버랩, 네이버지식인, 빅카인즈 각 25.0% 순이었다. 교수, 연구원은 다음/네이버카페, 국내일간지/온라인뉴스채널 42.9%, 복합데이터(뉴스+게시판+자료실+SNS) 28.6%, 구글트렌드, 다음SNS, 부동산114, 각종온라인게시판/댓글, 유튜브, 트위터, 인스타그램, 네이버블로그, 페이스북, 구글크롬로그파일(log file) 각 14.3% 순이었다.

민간 영역 빅데이터 사용 유형과 관련하여 주요 키워드는 다음과 같다.

카드매출 데이터, 이용자 분석, 핵심어, 온톨로지, 로봇크롤러, 여론동향조사, 댓글, 검색키워드, 유동인구 데이터, 생활인구, 영상

③ 민간 영역 빅데이터로 분석한 내용/주제 영역

민간 영역 빅데이터로 분석한 내용/주제 영역을 살펴보면 전체적으로 아동청소년 이외(성인 등) 내용/주제 영역이 45.5%로 가장 많았으며 다음으로 활동 및 문화 기타, 신체건강, 복지 및 권리 기타 각 36.40%, 중독, 공연/스포츠관람, 학교 폭력 27.30%, 학교생활, 정신건강 각 9.10% 순으로 나타났다. 전문가별로는 현장 전문가, 공무원의 경우 활동 및 문화 기타가 100.0%으로 가장 많았으며 다음으로 복지 및 권리 기타, 아동·청소년 이외(성인 등)가 각 25.0%였다. 교수, 연구원의 경우 신체건강이 57.1%로 가장 많았으며 다음으로 중독, 공연/스포츠관람, 복지 및 권리 기타, 비행 및 범죄 각 42.9%, 학교생활, 정신건강 각 14.3% 순이었다.

표 VI-8 민간 영역 빅데이터로 분석한 내용/주제 영역(다중응답)

구분			전문가				계	
			현장전문가, 공무원		교수, 연구원			
			빈도	%	빈도	%	빈도	%
민간 영역 빅데이터로 분석한 내용/주제 영역	교육 및 역량	학교생활	0	0.0	1	14.3	1	9.10
	건강 및 심리	신체건강	0	0.0	4	57.1	4	36.40
		정신건강	0	0.0	1	14.3	1	9.10
		중독	0	0.0	3	42.9	3	27.30
	활동 및 문화	공연/스포츠관람	0	0.0	3	42.9	3	27.30
		기타	4	100.0	0	0.0	4	36.40
	복지 및 권리	기타	1	25.0	3	42.9	4	36.40
	비행 및 범죄	학교폭력	0	0.0	3	42.9	3	27.30
	기타	아동청소년 이외 (성인 등)	1	25.0	4	57.1	5	45.50
	계			6	150	22	314.4	28

민간 영역 빅데이터로 분석한 내용/주제의 중요키워드는 다음과 같다.

청소년인권, 다음 포털 카페, 주제어, 대중교통 접근성, 소셜미디어, 감성분석, 주제도출, 정치여론, 청소년 자살위험 예측, 청(소)년 트렌드, 카드소비, 유동인구, 코로나 19, 학생, 학교, 교사, SNS, 라이프스타일, 사이버따돌림, 선거결과 예측, 학교폭력 유형, 아동청소년, 유튜브, 전파력, 유해가능성, 인터넷 중독 위험 예측, 메르스 정보확산 위험 요인 예측, 통일인식, 동향, 예측, 섹스팅 위험 예측, 담배 위험 예측, 머신러닝, 마약위험예측, 정보확산 위험 예측, 보건복지정책, 미래신호예측, 비만 미래신호예측

④ 민간 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항

민간 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항을 살펴보면 다음과 같다.

표 VI-9 민간 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항(다중응답)

구분	전문가				계		
	현장전문가, 공무원		교수, 연구원				
	빈도	%	빈도	%	빈도	%	
민간 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항 (다중응답)	개인정보보호로 인한 데이터활용, 결합의 한계	1	25.0	0	0.0	1	8.30
	공개데이터 양, 기간의 한계	1	25.0	2	25.0	3	25.00
	민간데이터 구매방법이 어렵거나 가격이 높음	3	75.0	1	12.5	4	33.30
	수집 및 분석에 큰 어려움이 없었음	0	0.0	3	37.5	3	25.00
	나에 대한 정보가 이미 많이 파악되고 있음에 대한 놀라움, 걱정	0	0.0	1	12.5	1	8.30
	연령 등 세부사항 특정하기 어려움	1	25.0	0	0.0	1	8.30
	분석의 한계(예: 트렌드분석은 가능하나 귀적연구는 어려운 것 등)	0	0.0	1	12.5	1	8.30
	공공데이터에 비해 데이터의 신뢰성, 타당성 확인 어려움	0	0.0	3	37.5	3	25.00
계	4	33.3	8	66.7	12	100.0	

전체적으로 민간데이터 구매방법이 어렵거나 가격이 높음이 33.3%로 가장 많았으며 다음으로 공개데이터 양, 기간의 한계, 수집 및 분석에 큰 어려움이 없었음, 공공데이터에 비해 데이터의 신뢰성, 타당성 확인 어려움이 각 25.0%, 개인정보 보호로 인한 데이터 활용, 결합의 한계, 나에 대한 정보가 이미 많이 파악되고 있음에 대한 놀라움, 걱정, 연령 등 세부사항 특정하기 어려움, 분석의 한계(예: 트렌드 분석은 가능하나 궤적 연구는 어려운 것 등)가 각 8.3%였다. 전문가별로는 현장전문가, 공무원의 경우 민간 데이터 구매 방법이 어렵거나 가격이 높음이 75.0%로 가장 많았으며 다음으로 개인 정보 보호로 인한 데이터 활용, 결합의 한계, 공개 데이터양, 기간의 한계, 연령 등 세부사항 특정하기 어려움이 각 25.0%였다. 교수, 연구원의 경우 수집 및 분석에 큰 어려움이 없었음, 공공데이터에 비해 데이터의 신뢰성, 타당성 확인 어려움이 각 37.5%였으며 다음으로 공개데이터 양, 기간의 한계 25.0%, 민간 데이터 구매방법이 어렵거나 가격이 높음, 나에 대한 정보가 이미 많이 파악되고 있음에 대한 놀라움, 걱정, 분석의 한계(예: 트렌드 분석은 가능하나 궤적 연구는 어려운 것 등)이 각 12.5%였다.

민간 영역 빅데이터 사용시 불편하거나 개선했으면 하는 사항의 중요의견은 다음과 같다.

- 네이더나 유튜브 사용시 알고리즘 때문에 놀란 적이 많음. 문자검색 외, 목소리도 인식하여 관련 사이트나 동영상을 자동 연결하는 것을 보면 구글이나 네이버가 나에 대해 너무 많은 것을 알고 있다는 생각이 들어 놀랍기도 하고 우려도 됨.
- 종단적 자료를 수집해도 패널 자료가 아니어서 트렌드 분석은 가능하나 궤적 연구에는 어려움이 있음.
- 분석의 대상이 사람이 아닌 문서이다 보니 일반적으로 통용되는 사회과학 연구 질문을 검증하고 이해하기 쉽고 의미 있게 함의를 도출하는데 어려움이 있음.

3. 빅데이터 관련 욕구

1) 빅데이터에 대한 관심 및 활용 희망 여부

(1) 빅데이터에 대한 관심

빅데이터에 대한 관심에 대해 살펴본 결과 전체적으로 보통 10.8%, 관심있다 35.1%, 매우 관심 있다 54.1%로 나타났으며 평균은 4.4점(백분위 점수 85.7점)에 해당하였다. 전문가별로는 현장전문가, 공무원은 보통 11.1%, 관심있다 33.3%, 매우 관심있다 55.6%으로 평균은 4.4점(백분위 점수 86.0점), 교수, 연구원은 보통 10.7%, 관심있다 35.7%, 매우 관심있다 53.6%로 평균 4.4점(백분위 점수 85.7점)으로 나타나 빅데이터에 대한 관심이 높음을 알 수 있다.

표 VI-10 빅데이터에 대한 관심

구분		전문가				계	
		현장전문가, 공무원 (N=9)		교수, 연구원 (N=28)			
		빈도	%	빈도	%	빈도	%
빅데이터에 대한 관심	전혀 관심 없다 (1점)	0	0.0	0	0.0	0	0.0
	별로 관심 없다 (2점)	0	0.0	0	0.0	0	0.0
	보통 (3점)	1	11.1	3	10.7	4	10.8
	관심 있다 (4점)	3	33.3	10	35.7	13	35.1
	매우 관심 있다 (5점)	5	55.6	15	53.6	20	54.1
계		9	100.0	28	100.0	37	100.0
χ^2		0.0					
(평균, 백분위 점수, SD)		(4.4, 86.0, 0.7)		(4.4, 85.7, 0.6)		(4.4, 85.7, 0.6)	
(최소, 최대)		(3, 5)		(3, 5)		(3, 5)	
t 값		.05					

*p<.05, ** p<.01, *** p<.001

(2) 1년 이내 빅데이터 활용 희망 여부

“1년 이내에 빅데이터를 활용하여 연구 또는 분석을 할 생각”에 대해 질문한 결과 전체적으로 83.8%가 있다고 응답하였으며 전문가별로는 현장전문가, 공무원은 77.8%, 교수, 연구원은 85.7%가 그렇다고 응답하였다.

표 VI-11 1년 이내 빅데이터 활용 희망 여부

구분		전문가				계	
		현장전문가, 공무원		교수, 연구원			
		빈도	%	빈도	%	빈도	%
1년 이내 빅데이터 활용 희망 여부	있음	7	77.8	24	85.7	31	83.8
	없음	2	22.2	4	14.3	6	16.2
계		9	100.0	28	100.0	37	100.0
χ^2		0.3					

*p<.05, ** p<.01, *** p<.001

2) 빅데이터 활용 희망하지 않는 경우 이유⁸⁰⁾

1년 이내에 빅데이터 활용을 희망하지 않는 경우 그 이유에 대해 질문한 결과 전체적으로 시간적 제약 또는 다른 방법론에 관심을 우선 두고 있어서, 빅데이터의 신뢰성 및 타당성, 학문적 깊이 부족, 체계적 관리 미흡, 오류 응답 가능성으로 인해가 42.9%로 가장 많았으며 다음으로 분석할 수 있는 데이터와 데이터 접근방법에 대한 정보가 부족하여 28.6%, 빅데이터를 탐색하고 연구하는 방법론을 학습하는 시간이 필요하여 14.3% 순이었다.

80) 추후 활용하고 싶은 데이터에 대한 내용은 앞에서 살펴본 사용해본 빅데이터 유형과 대동소이하여 별도로 따로 제시하지 않았다.

표 VI-12 빅데이터 활용 희망하지 않는 경우 이유

구분		전문가구분R				계	
		현장전문가, 공무원		교수, 연구원			
		빈도	%	빈도	%	빈도	%
빅데이터 활용 희망하지 않는 경우 이유	시간적 제약 또는 다른 방법론에 관심을 우선 두고 있어서	0	0.0	3	60.0	3	42.9
	분석할 수 있는 데이터와 데이터 접근방법에 대한 정보가 부족하여	2	100.0	0	0.0	2	28.6
	빅데이터를 탐색하고 연구하는 방법론을 학습하는 시간이 필요하여	0	0.0	1	20.0	1	14.30
	빅데이터의 신뢰성 및 타당성, 학문적 깊이 부족, 체계적 관리 미흡, 오류 응답 가능성으로 인해	0	0.0	3	60.0	3	42.9
계		2	100.0	7	140.0	9	128.7

전문가별로 현장전문가, 공무원은 분석할 수 있는 데이터와 데이터 접근방법에 대한 정보가 부족하여가 100.0%로 전체를 차지하였으며 교수, 연구원은 시간적 제약 또는 다른 방법론에 관심을 우선 두고 있어서, 빅데이터의 신뢰성 및 타당성, 학문적 깊이 부족, 체계적 관리 미흡, 오류응답 가능성으로 인해 각 60.0%, 빅데이터를 탐색하고 연구하는 방법론을 학습하는 시간이 필요하여 20.0% 순으로 나타났다.

○ ————— 제7장 결론 및 제언

- 1. 청소년 빅데이터 체계 구축을 위한 비전 및 전략
- 2. 청소년 빅데이터 체계 구축 및 활용을 위한 세부 정책과제 도출
- 3. 청소년 빅데이터 온톨로지(안) 제시

1. 청소년 빅데이터 체계 구축을 위한 비전 및 전략⁸¹⁾

연구내용을 바탕으로 청소년 빅데이터 체계 구축을 위한 비전 및 전략을 제시하면 다음과 같다.



그림 VII-1 청소년 빅데이터 체계 구축을 위한 비전 및 전략

81) '제7장 결론 및 제언'은 한국청소년정책연구원 서정아 선임연구위원이 작성하였다.

현재 아동·청소년과 관련하여 가장 공신력 있는 데이터를 다수 확보하고 있는 중앙행정기관은 ‘통계청’이라고 할 수 있다. 그러나 통계청 데이터의 경우 단순한 기술치를 제시하는 방식으로 자료가 제시되어 있어 연구자, 정책가 등의 다양한 활용이 쉽지 않은 점이 있다. 전체 데이터 셋을 확인하기 어려울 뿐 아니라 각 지표의 정의, 측정대상(연령 등), 측정방법 또한 상이하여 비교분석이 쉽지 않을 뿐 아니라 데이터를 활용 가능한 형태로 만드는 데에도 상당한 수고와 노력이 뒤따른다. 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 공신력 있는 자체 자료를 다수 확보하여 공개하고 있으며 이미 상당수의 데이터들이 축적되어 있어 비교적 빅데이터 허브로서의 역할로 전환하는 데 큰 무리가 없는 것으로 보인다. 예를 들어 한국청소년정책연구원의 한국 아동·청소년 데이터 아카이브의 경우 청소년 분야의 유일한 국책연구기관으로 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 가능하다. 이에 기존의 한국아동·청소년 데이터 아카이브를 ‘(가칭) 한국아동청소년빅데이터플랫폼’으로 확대하여 아동·청소년 분야 데이터 허브로 활용하는 것을 제안한다. 이 과정에서 1단계로 통계청의 국가통계포털, 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 보인다. 2단계로는 지방자치단체, 공공기관(한국청소년활동진흥원 빅데이터 허브 등), 대학의 데이터 허브로 데이터 공유, 연계 체계를 확대하고 3단계로 국제기구, 기업빅데이터 플랫폼, 비정형 빅데이터 등으로의 확대를 제안한다.

2. 청소년 빅데이터 체계 구축 및 활용을 위한 세부 정책과제 도출

청소년 빅데이터 체계 구축 및 활용을 위한 세부 정책과제를 제안하면 다음과 같다.

1) 빅데이터 활용 향상을 위한 세부 정책과제

연구결과를 바탕으로 빅데이터 활용 향상을 위한 세부 정책과제를 제시하면 다음과 같다.

- 4차 산업혁명의 기초자원이자 결정적 요인으로써 빅데이터의 개념, 활용에 대한 인식 제고
- 빅데이터에 대한 이해를 제고하고 활용성을 향상하기 위한 다양한 교육 기회 제공
- 다양한 빅데이터 분석방법론(다변량 분석, 구조방정식, 미래신호탐색, 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
- 공공기관의 공공데이터와 공신력을 어느 정도 갖추고 있는 민간 기관의 데이터를 한 곳에서 볼 수 있는 빅데이터 플랫폼, 중앙아카이브 시스템을 구축, 운영
- 데이터의 책임 있는 생산과 관리 수준 향상, 전문성 강화
- 빅데이터 연구자간의 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

2) 미래예측을 위해 우선적으로 구축되어야 할 청소년 관련 빅데이터 세부 영역

미래예측을 위해 우선적으로 구축되어야 할 청소년 관련 빅데이터 세부 영역을 제시하면 다음과 같다.

- 복지사각지대, 범죄 등이 발생할 수 있는 부분
- 안전사고 사전 예방 관련 부문
- 아동·청소년의 목소리를 들을 수 있는 데이터
- 청소년의 소비, 경제활동 데이터
- 유아기부터 성인기까지 전생애적 성장과정(개인, 가족, 학교, 또래 등)을 추적할 수 있는 데이터
- 청소년정책효과 추정가능 마이크로 자료
- 청소년정책 효과성에 대한 인지 수준
- 청소년의 일상생활, 라이프스타일, 생활패턴, 소비패턴 등에 관한 데이터(교통카드, 청소년증 등)

- 청소년들의 삶에 대한 생각, 태도, 가치관
- 진로행동, 진로의식, 직업선호도, 취업/창업 등 진로관련 데이터
- 환경에 대한 인식
- 시민권 및 정치권 관련 데이터
- 청소년 정서발달
- 청소년전담(담당) 공무원 현황 등
- 국제적 표준, 국제적 비교 등이 가능하도록 하는 국제적 공통지표, 항목을 포함한 데이터(예: 국제연합(UN)의 지속가능개발 목표 포함 등)

3) 아동·청소년 분야 빅데이터 체계 구축을 위한 세부 정책과제

아동·청소년 분야 빅데이터 체계 구축을 위한 세부 정책과제를 제안하면 다음과 같다.

- 아동·청소년을 바라보는 시각을 조정: 전반적으로 통제와 보호의 대상으로 아동·청소년을 바라보는 시각에서 시민권을 가진 인격적 주체로 바라보는 시각으로 전환하여 지표체계를 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. 현재로써는 행정조직 특유의 부서주의로 인해 정보의 공유나 제공이 쉽지 않음. 문제를 해결하기 위해서는 이러한 빅데이터를 여러 정부부처나 기관으로부터 제공 받아 정제하고 분석할 뿐 아니라 연구자들과 일반인들에게 용도에 맞게 차등화된 수준의 데이터를 공급하는 역할을 수행할 일종의 빅데이터통합센터의 역할을 할 기관이 필요함.
- 증거기반 정책, 서비스 개발 및 제공에 대한 인식 강화. 데이터 기반 문제해결 프로세스 운영을 위한 조직구성, 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 수집해 주는 것이 향후 효과적인 데이터 이용과 문제해결을 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 민관 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
- 다양한 교육, 연수 기회 제공
- 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 데이터범위 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, TicToc 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
- 아동·청소년 분야의 빅데이터가 효과적으로 구축되기 위해서는 어느 한 기관이 데이터를 수집 관리하기 보다는 플랫폼을 만들고 각 기관에서 생산되는 여러 가지 데이터를 연계하여 분석할 수 있도록 하는 시스템을 마련하는 것이 보다 효과적일 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
- 사례데이터의 체계적 수집과 활용을 위한 전자문서 서식화 필요
- 데이터 생산과 관리의 책무성에 대한 인식 강화 필요

3. 청소년 빅데이터 온톨로지(안) 제시

마지막으로 청소년 분야 빅데이터 온톨로지(안)과 관련하여 [표 III-2]에 추가되어야 하는 사항을 중심으로 제시하면 다음과 같다.

- **인구 및 가족 분야:** 인구이동, 결혼/출산 인식, 다문화 아동·청소년수, 중도입국청소년수, 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 함께하는 활동, 대화시간, 함께 식사하는 횟수, 주거형태, 코로나19 이후 생활 변화 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·청소년돌봄형태, 학교유형, 학습시간, 진로 및 직업관련 역량, 방과후활동 참여여부, 미디어 및 SNS 활용관련 역량, 또래간 폭력, 언어적 역량, 교내 단체활동 참여여부, 문화예술능력, 문화감수성, 사교육 분야, 리더십
- **건강 및 심리 분야:** 공동체의식(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), 코로나 19 등 재난 이후 건강 및 심리 분야 변화, 자기이해(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, 미디어 리터러시, 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 진로관여, 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, 문화활동 중 문화예술 분야에서 점차 확대되는 문화예술 교육 참여 경험, 참여율, 프로그램, 예술활동, 청소년활동 참여동기, 청소년참여기구 현황, 여가활동 유형, 인증수련활동 현황, 청소년활동종사자, 코로나19이후 청소년활동 및 문화생활 경험 등
- **복지 및 권리/참여 분야:** 정치 참여, 의식 및 가치(시민의식 등), 이주배경 청소년, 인터넷, 게임, 도박중독청소년, 청소년자립지원시설 이용, 청소년참여 활동여부(청소년운영위원회, 청소년특별회의, 참여위원회), 청소년 노동

권, 청소년증 현황, 청소년증 사용현황, 청소년 액티비티, 문화활동, 학교 밖 청소년 여부, 청소년 참여에 대한 인식, 가정에서의 청소년권리 보장정도, 가정내 주요의사결정 참여 정도 등

- **일상생활분야:** 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

참고문헌

[국내 문헌]

- 강희정(2015). 보건의료 빅데이터 활용을 위한 기본계획 수립. 세종: 한국보건사회연구원.
- 고수정(2018). 빅데이터 군집 분석을 이용한 학습성취도 예측: 종단연구를 중심으로. 한국디지털콘텐츠학회, 19(9), 1,769-1,778.
- 권영옥, 빅데이터를 활용한 맞춤형 교육서비스 활성화 방안 연구. 지능정보연구, 19, 2, 87-100.
- 권정은, 정지선(2012). 소셜 분석으로 살펴본 청소년 자살예방정책의 시사점. 서울: 한국정보화진흥원 빅데이터전략연구센터.
- 김기현, 김창환(2011). 한국 아동·청소년 종합통계체계 구축 연구. 서울: 한국청소년정책연구원.
- 김수연, 이수진, 이나라, 김진주, 김민아, 김지윤, 양윤나, 이자람, 한은지, 김우찬, 허아람, 한승희, 안소영, 이도형(2018). 혁신성장동력분야 주요 경쟁국 정책 모니터링 및 심층분석. 한국과학기술기획평가원.
- 김지은, 백순근(2016). 텍스트 빅데이터 분석 기법을 활용한 대학구조개혁 평가의 쟁점 분석. 서울대학교 교육연구소. 아시아교육연구 제17권 제 3호.
- 김형주, 최인재, 임지연(2008). 한국 청소년 발달지표 조사 III: 청소년종합통계체계 구축방안 연구(연구보고 08-R03-2). 서울: 한국청소년정책연구원.
- 노만 매트로프 지음, 권정민 역(2016). 데이터 고급 분석과 통계 프로그래밍을 위한 빅데이터 분석 도구 R 프로그래밍. 서울: 갑우문화사.

- 노법래. (2015). 자유세션 1: 아동/청소년; 증거기반 사회정책 수립을 위한 빅데이터 (big data) 와 지리정보시스템 (GIS)의 정책 수요 분석에의 적용: 청소년 자살과 정신건강 서비스 사례를 중심으로. 한국사회보장학회 정기학술발표논문집, 2015(2), 1,033-1,051.
- 문유경, 전기택, 김영택, 정성미, 배호중, 정희태, 김유정(2018). 여성가족 빅데이터의 활용을 통한 정책 개선과 발굴(II). 서울: 한국여성정책연구원.
- 박준석(2019). 빅 데이터 등 새로운 데이터에 대한 지적재산권법 차원의 보호가능성. 산업재산권, 58, 77-129.
- 박찬국, 김현제(2015). 사물인터넷을 통한 에너지 신산업 발전 방향 연구-텍스트 마이닝을 이용한 미래 신호 탐색. 울산광역시: 에너지경제연구원.
- 배지영, 조미숙(2016). 소셜빅데이터 분석을 위한 식품안전 온톨로지 구축. Journal of Health Informatics and Statistics. 278-289.
- 백지연, 김태영, 양동민, 오효정(2018). 빅데이터 기반 도서관 어린이, 청소년서비스 현황분석 및 개선방안. 한국도서관정보학회지, 49(4), 295-328.
- 백혜정, 송미경, 신정민(2015). 학교 밖 청소년지원정책 체계화 방안 연구. 세종: 한국청소년정책연구원.
- 벤 웨이버 저, 배출효 역(2015). 구글은 빅데이터를 어떻게 활용했는가. 서울: 북카라반.
- 손해인(2019). 자살예방의 도구로 활용되는 뉴욕주 빅데이터 - PSYCKES. 국제사회보장리뷰, 10 가을호, 114-126.
- 송영조(2012). 빅데이터 시대! SNS의 진화와 공공정책. 서울: 한국정보화진흥원.
- 송주영, 송태민(2014). 소셜 빅데이터를 활용한 북한 관련 위협인식 요인 예측. 국제문제연구, 가을, 209-243.
- 송태민(2013). 소셜 빅 데이터분석을 통한 자살 검색 예측모형 개발. 보건복지포럼, 202, 74-86.

- 송태민(2013). 우리나라 보건복지 빅데이터 동향 및 활용방안. 과학기술정책, 192, 56-73.
- 송태민(2015). 소셜 빅데이터를 활용한 우리나라 청소년의 우울 현황과 위험요인 분석. 보건복지 Issue & Focus, 296.
- 송태민, 송주영, 진달래(2014). 소셜 빅데이터를 활용한 인터넷 중독 위험예측 모형. 보건사회연구, 34(3), 106-134.
- 송태민, 송주영(2015). 빅데이터 연구 한 권으로 끝내기. 한나래아카데미.
- 송태민, 송주영(2016). R을 활용한 소셜 빅데이터 연구방법론. 서울: 한나래아카데미.
- 송태민, 송주영(2017). 머신러닝을 활용한 소셜 빅데이터 분석과 미래신호 예측.
- 송태민, 송주영, 천미경(2016). 소셜 빅데이터와 Google 검색 트렌드를 활용한 한국과 미국의 사이버불링 검색에 영향을 미치는 요인 분석. 한국빅데이터어학 회, 1(1), 67-75.
- 송태민, 이기호, 진달래, 천미경, 서동철, 박현애(2016). 2016년 소셜 빅데이터 기반 보건복지 이슈 동향 분석. 세종: 한국보건사회연구원.
- 송태민, 이중순(2014). 일본의 빅데이터 프라이버시 보호방안. 보건복지포럼, 210, 89-99.
- 송태민, 진달래, 박대순, 박현애, 안지영, 김정선(2014). 보건복지 빅데이터 효율 적 관리방안 연구. 세종: 한국보건사회연구원.
- 안명숙(2018). 신조어 청소년 혐오에 관한 소셜 미디어 빅데이터 분석. 한국청소년시설통계학회, 16(1), 통권 제55호, 167-176.
- 안형준(2016). 미국 알고리즘 안에 내재된 사회적 차별: 빅데이터에 대한 미국 정부의 우려. 과학기술정책, 26(5), 4-7.
- 여성가족부(2019). 2019 청소년백서. 서울: 여성가족부.
- 유민상, 유성렬, 이희길, 정지은, 도남희, 구자연, 류정희, 김혜자(2019). 한국 아동·청소년 삶의 질 지표 구축 방안 연구. 세종: 경제·인문사회연구회 협동연구총서.
- 윤은경, 박상희(2019). 빅데이터에 나타난 청소년 생애 설계에 대한 의미 분석. Asia-pacific Journal of Multimedia Services Convergent with Art,

- Humanities, and Sociology, Vol. 9, No.3, March(2019), pp.1-8.
- 윤진수, 고광표, 박규리, 김영균, 이효철, 테레시아, 이석원(2017). 빅데이터를 통한 청소년 폭력 예방 시스템. 한국정보과학회 학술발표논문집, 1,944 - 1,946.
- 윤창근(2012), 미국: IBM 왓슨(Watson)을 통해 살펴 본 인공지능 기술과 빅데이터의 접목과 활용가능성. Local Information Magazine. 2012, May+June. 74. 68-73.
- 이경상, 이창호, 김 민(2019). 지능정보사회에서의 청소년활동정책 개선방안: 프로그램, 설비·시설, 지도자를 중심으로. 정보화정책, 26, 4, 62-84.
- 이상엽(2015). 미국 NSF, 미국 전역에 대한 빅데이터 허브 구축계획 발표. 과학기술정책 25(11), 4-5.
- 이은경, 박도영, 최인봉(2014). 교육 빅데이터 관련 연구동향. 한국컴퓨터정보학회 하계학술대회 논문집, 22, 2, 175-176.
- 이지혜, 제미경, 조명지, 손현석(2014). 보건의료분야의 빅데이터 활용동향. 한국통신학회지(정보와 통신), 12, 63-75.
- 이현웅, 김종엽, 최현재(2018). 빅데이터의 이해와 활용: 공공 부문을 중심으로. 서울: 생각나눔.
- 임희준(2020). 키워드 네트워크 분석을 통해 살펴본 초등학생이 인식하는 과학 학습 참여의 의미, 초등과학교육, 39(2), 255-267
- 정보통신기술진흥센터(2016). 미국의 빅데이터 R&D 전략계획. 해외ICT R&D 정책동향. 2016-12호. 1-35. 서울: 정보통신기술진흥센터.
- 정혜실(2016). 청소년우울 관련 소셜데이터 수집과 분석을 위한 온톨로지 개발 및 평가, 서울대학교 석사학위 논문.
- 조주연, 김현숙, 조민재(2018). 텍스트마이닝을 활용한 청소년 문제의 이슈 변화 분석 - 2008년~2018년 인터넷 신문기사를 중심으로. 교육혁신연구 28(4), 461-487.
- 최용환, 성윤숙, 박상현(2018). 증거기반 정책 수립을 위한 아동청소년청년 통계 관리체계화방안. 세종: 한국청소년정책연구원.

- 한국데이터산업진흥원(2019). 2019 데이터산업 백서.
- 한국정보화진흥원(2012). 소셜분석으로 살펴본 청소년자살예방정책의 시사점.
- 한국정보화진흥원(2019). 2019 국가정보화백서.
- 한윤선, 김하영, 송주영, 송태민(2019). 소셜 빅데이터 수집 및 분석을 위한 아동 청소년 학교폭력 온톨로지 개발. 한국콘텐츠학회, 19(6), 10-23.
- 황수정(2019). 미·중·일의 과학기술정책 및 연구개발 동향 비교, 한국연구재단 R& D BRIEF, 2019-37호(2019.08.26.).
- 후나오 노부오 저, 김성재 역(2014). R로 배우는 데이터 분석 기본기 데이터 시각화. 서울: 한빛미디어.

[외국 문헌]

- ADR UK (2019). Administrative data is an invaluable resource for public good. Let's use it (Annual Report 2018-2019). Swindon, Wiltshire: ADR UK.
- Bachner, J. (2013). Predictive policing: Preventing crime with data and analytics. Washington, D.C. IBM Center for the Business of Government.
- Connelly, R., Playford, C. J.m Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. Social Science Research, 59, 1-12.
- De Angelis, J. & Perasso, G. (2020). Cyberbullying detection through machine learning: Can technology help to prevent internet bullying? International Journal of Management and Humanities. 4(11), 57-66.
- Department for Business, Innovation & Skills (2013). Seizing the Data Opportunity: A Strategy for UK Data Capability. London: Department

- for Business, Innovation & Skills.
- Department for Business, Innovation & Skills (2014). Open Data Strategy 2014-2016. London: Department for Business, Innovation & Skills.
- Department for Digital, *Culture, Media & Sport* (2017). UK Digital Strategy.
- Department for Digital, Culture, Media & Sport (2017). UK Digital Strategy. London: Department for Digital, Culture, Media & Sport.
- Federal Interagency Forum on Child and Family Statistics(2019). America's Children: Key National Indicators of Well-Being, Washington, DC: U.S. Government Printing Office.
- 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?.
- Giusine, D, Fraboni, F., De Angelis, M., & Pietrantonio, L. (2019). Commentary: Princiles, approaches and challenges of applying big data in safety psychology research. *Frontiers in Psychology*, 10, 1-3.
- Government Digital Service (2017). Government Transformation Strategy. London: Government Digital Service.
- Guntuku, S.C., Yaden, D., Kern, M.L., Ungar, L.H., & Eichstaedt, J.C.(2017). Detecting depression and mental illness on social media: An integrative review. *Current Opinion in Behavioral Sciences*, 18,

43-49.

- Harari, G. M, Muller, S. R., Aung, M. S, & Rentfrow, P. J. (2017). Smartphone sensing methods for studying behavior in everyday life. *Current Opinion in Behavioral Sciences*, 18, 83-90.
- Information Commissioner's Office (2017). Information Rights Strategic Plan 2017-2021. Wilmslow, Cheshire: International Commissioner's Office.
- 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.
- Leiter, M. (2013). Crime modeling and mapping using geospatial technologies. Heidelberg: Springer.
- Manyika, J & Chui, M. (2011). Big data: The next frontier for innovation, competition, and productivity, McKinsey Global Institute.
- OECD(2011), 15th Meeting of the working party on indicators for the information society.
- Roberts, S. (2019). How roundtables are helping inform the National Data Strategy. DCMS Blog(<http://dcmsblog.uk>). Retrieved from
- Rumelhart DE, GHinton GE, Williams RJ.(1986). Learning representations by back-propagating errors. *Nature*, 323, 533-536.
- Snijders, C., Matzat, U., & Reips, U.-D. (2012). 'Big Data': Big gaps of knowledge in the field of Internet. *International Journal of Internet Science*, 7, 1-5.
- The World Bank (2014). Open Data for Economic Growth. Washington, D.C.: The World Bank.
- Thompson, K. (2011). The Santa Cruz Experiment: Can a City's Crime

- Be Predicted and Prevented?' Popular Science, 11. 2.
- UNICEF, I. (2007). An overview of child well-being in rich countries. Innocenti Report Card, 7(7), 34.
- United Nations (2020). United Nations E-Government Survey 2020. New York, NY: UN.
- 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.
- White House Office of Science and Technology Policy(2020). American artificial intelligence initiative: Year one annual report.
- White House(2016). Big data: A report on algorithmic systems, opportunity, and civil rights. Executive Office of the President.
- 노무라연구소 小林慎太郎(2012. 11). 빅데이터사회におけるプライバシー~「個人情報」から「プライバシー」の保護へ~ 第176回 NRIメディアフォーラム資料.
- 鈴木 正朝(2012. 11.). プライバシーの権利と個人情報保護法. 総務省マイナンバーシンポジウム資料.
- 中田響(2007). 個人情報性の判講造. 慶藤義塾大メディア・コミュニケーション研究所紀要, 57, 145-161.

[웹사이트 및 홈페이지 자료]

https://commons.wikimedia.org/wiki/File:Artificial_neural_network.png

<https://kbig.kr/portal/kbig>

<http://www.bigdata.go.kr>

<https://www.kihasa.re.kr/>

<https://www.kipa.re.kr/site/kipa/main.do>

<https://www.data.go.kr>
<http://agict.gov.au/policy-guides-procurement/cloud>
<http://agimo.gov.au/2013/03/15/released-big-data-strategy-issues-paper/>
http://agimo.gov.au/ict_strategy_2012_2015.
<http://agimo.gov.au/policy-guides-procurement/cloud/>
<http://dataavailability.pmc.gov.au/sites/default/files/govt-response-pc-dau-inquiry.pdf>
<http://gov.uk>
<http://ico.org.uk>
<http://nebigdatahub.org/nsdc-seedfund/>
<http://theodi.org>
<http://webguide.gov.au/web-2-0/publishing-public-sector-information/>
<http://www.itdaily.kr/news/articleView.html?idxno=100244>
<http://www.korea.kr/special/policyCurationView.do?newsId=148867915>
<http://www.ndsl.kr/ndsl/issueNdsl/detail.do?techSq=121>
<http://www.nss.gov.au/dataintegration>
<http://www.nss.gov.au/nss/home.nsf/NSS/7AFDD165E21F34FDCA2577E400195826?opendocument>
<http://www.nss.gov.au/nss/home.NSF/pages/High+Level+Principles+for+Data+Integration+-+Content?OpenDocument>
<http://www.nss.gov.au/nss/home.NSF/pages/Statistical%20Spatial%20Framework>
<http://www.oaic.gov.au/privacy/privacy-engaging-with-you/previous-privacy-consultations/de-identification-resources-may-2013/>
<http://www.oaic.gov.au/privacy/privacy-resources/privacy-guides/privacy-impact-assessment-guide>

<https://bigdata.busan.go.kr/index.busan>
<https://builtin.com/big-data/big-data-platform>
<https://catalog.data.gov/>
<https://datacenter.kidscount.org/>
<https://datacenter.kidscount.org/>
<https://datacenter.kidscount.org/characteristics>
<https://datacenter.kidscount.org/topics>
https://en.wikipedia.org/wiki/Health_Insurance_Portability_and_Accountability_Act
<https://gov.uk>
<https://healthdata.gov/>
<https://healthdata.gov/>
https://itlaw.wikia.org/wiki/Networking_and_Information_Technology_Research_and_Development_Program
<https://kosis.kr>
<https://nebigdatahub.org/news-updates>
https://omh.ny.gov/omhweb/psyckes_medicaid/about
<https://openpolicy.blog.gov.uk>
<https://toolkit.data.gov.au/Planning.html>
<https://uksa.statisticsauthority.gov.uk>
<https://www.adruk.org>
<https://www.aecf.org/resources/2020-kids-count-data-book>
<https://www.aihw.gov.au/reports/children-youth/scoping-enhanced-measurement-of-child-wellbeing/contents/table-of-contents>
<https://www.bigdatahubs.io/>
<https://www.childstats.gov/>

https://www.childstats.gov/americaschildren/data_topics.asp
<https://www.childstats.gov/americaschildren/members.asp>
<https://www.childstats.gov/pubs/>
<https://www.childtrends.org/>
<https://www.childtrends.org/blog/big-data-and-social-science-research>
<https://www.childtrends.org/indicators?a-z>
<https://www.congress.gov/>
<https://www.congress.gov/115/plaws/publ435/PLAW-115publ435.pdf>
<https://www.congress.gov/bill/115th-congress/house-bill/4174/related-bills>
<https://www.data.go.kr/guide/guide/supportIntroduce.do>
<https://www.data.gov.au/>
<https://www.data.gov/>
<https://www.data.gov/about>
<https://www.data.gov/meta/data-gov-at-ten-and-the-open-government-data-act/>
<https://www.data.gov/meta/data-gov-at-ten-and-the-open-government-data-act/>
<https://www.data.gov/ocean/>
<https://www.dsta.gov.sg/docs/default-source/dsta-about/risk-assessment-and-horizon-scanning-experimentation-centre.pdf?sfvrsn=2>
<https://www.gartner.com>
<https://www.gov.uk/government/groups/horizon-scanning-centre>
<https://www.icas.com/thought-leadership/technology/10-companies-using-big-data>
<https://www.industry.gov.au/sites/default/files/2018-12/data-strategy>

<https://www.informationweek.com/big-data/ai-machine-learning/trump-white-house-launches-ai-initiative/d/d-id/1333845>

https://www.kywa.or.kr/management/public_10_01.jsp

<https://www.nhis.or.kr/retrieveHomeMain.xx>

https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505185

<https://www.nypi.re.kr/archive/contents/siteMain.do>

<https://www.oaic.gov.au/>

<https://www.premiers.qld.gov.au/publications/categories/plans/open-data-strategy.aspx>

https://www.privacy.go.kr/pic/nation_usa.do

https://www.privacy.go.kr/pic/nation_usa05.do

<https://www.stopbullying.gov/>

<https://www.stopbullying.gov/>

<https://www.whitehouse.gov/ai/>

<https://www.whitehouse.gov/wp-content/uploads/2018/02/ap-18-research-fy2019.pdf>

https://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf

<https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html>

<https://youth.gov/>

<https://youth.gov/federal-resources>

<https://unstats.un.org/bigdata/inventory/?selectTitle=&selectCountry=United+States&selectOrganization=U.S.+Office+of+Management+and+Budget&selectSource=&selectstatArea=&selectsdgGoal=>

부 록

- 1. 전문가 심층면접지
- 2. 전문가 설문조사지
- 3. 아동청소년관련 승인통계
변인 조사표(예)

부록

1. 전문가 심층면접지

아동·청소년 빅데이터 체계 구축 및 활용방안 연구 심층 면접조사 설문지

○ 조사목적 및 안내

- 이 조사는 아동·청소년 빅데이터 체계 구축 및 활용방안에 대하여 관련 업무 수행 경험을 가진 청소년정책 및 현장 관련 학계, 행정계, 현장 전문가를 대상으로 인터뷰 형태로 진행되는 심층면접 조사입니다.
- 응답한 내용 및 응답자의 개인 정보와 관련된 부분은 모두 비밀이 보장됩니다. 다만 조사의 내용 분석을 위해 녹음이 필요합니다. 이에 대한 양해 부탁드립니다.
- 인터뷰 진행은 약 1시간 ~ 1시간 30분 내외로 진행될 예정입니다.
- 연구에 참여해주셔서 감사합니다.

문2 본 연구에서 청소년 빅데이터를 ‘청소년의 다양한 삶의 영역에서 다양한 형태 (Variety)로 빠르게(Velocity) 생성되는, 하지만 휘발성(Volatility)을 지닌 대규모(Volume)데이터 중 유의미한 데이터의 선별, 취합 및 생산적 분석을 통해, 사회가 직면하고 있는, 또는 직면할 수 있는 다양한 청소년 문제의 적시성 있고 적절한 미래지향적 해소가 가능하도록, 가변성(Variability)을 고려하여 정확성(Veracity)을 지닌 가치(Value)로서의 생산적 정보를 도출 및 시각화(Visualization)할 수 있는 기술 및 능력’으로 정의하였습니다.

청소년 빅데이터의 개념에 대한 수정 및 추가 의견 있으시면 말씀해 주세요?

문3 3-1) 국내외의 빅데이터 체계 구축 상황과 우수사례에 대해 아는 범위 내에서 말씀해 주세요.

3-2) 청소년 빅데이터 체계 구축 및 활용의 구체적인 우수 사례를 말씀해주세요?

- 국내 사례 :
- 국외 사례 :

문4 4-1) 빅데이터를 활용해 본 경험이 있으신 경우 경험에 대해 말씀해 주세요.

4-(1) 무엇을 활용하셨는지?

4-(2) 왜 활용하셨는지?

4-(3) 어떻게 활용하셨는지?

4-2) 만약 아직까지 활용해보신 경험이 없다면 어떠한 데이터를 활용해서 어떠한 사항을 분석해보고 싶으신지 의견을 자유롭게 말씀해 주세요

문5 현재 활용가능한 정형, 비정형 빅데이터로 알고 계신 것을 자유롭게 말씀해 주세요.

문6 청소년 대상 빅데이터에 어떠한 것이 있는지 분류하는 지표 체계를 중분류로 영역화하여 세부항목을 산정한다고 할 경우, 청소년과 관련하여 꼭 들어가야 한다고 생각하는 중분류, 세부항목에는 어떠한 것이 있나요?

- 중분류: 예시) 인구, 가족, 교육, 역량, 건강, 안전, 심리, 활동, 주관적 복지감 등
- 세부항목: 예시) 방과후 나홀로 보내는 시간, 독서율, 비만도

문7 청소년과 관련된 빅데이터 체계를 만들 때 꼭 들어가야 할 세부지표, 알고 싶은 것은 무엇입니까?

문8 청소년복지 증진 및 정책개발, 미래예측을 위해 청소년 빅데이터를 통해 우선적으로 분석, 도출되어야 하는 인과관계(독립변인, 종속변인)에는 무엇이 있을까요? 중요하다고 생각하는 사항을 말씀해 주세요.

문9 청소년 통계활용에서 정형, 비정형 빅데이터의 활용도를 높이기 위한 방안은 무엇이라고 생각하십니까?

문10 현재 있는 다양한 청소년관련 빅데이터를 어떻게 체계화 하여 활용할 수 있도록 하면 이용자 입장에서 편리할까요?

문11 청소년 빅데이터 체계 구축 및 활용을 하기 위한 대안으로 청소년관련 빅데이터를 관할하는 컨트롤 타워를 만들어서 자료의 활용성을 높이는 것에 대해 어떻게 생각하십니까?

11-1) 컨트롤 타워가 필요하다면 그 이유가 무엇이고, 어떤 부처와 기관에서 컨트롤 타워 역할을 해야 한다고 보십니까?

11-2) 컨트롤 타워가 필요하지 않다면 그 이유가 무엇입니까?

문12 청소년 빅데이터 체계 구축 및 활용의 장점은 무엇이라고 생각하십니까?

문13 청소년 빅데이터 체계 구축 및 활용의 단점은 무엇이라고 생각하십니까?

문14 한국의 상황에서 청소년 빅데이터 체계 구축 및 활용의 한계점은 무엇이라고 생각하십니까?

문15 빅데이터를 잘 활용할 수 있도록 하기 위해 어떠한 부분이 개선되면 될까요?

문16 청소년정책에 빅데이터를 어떻게 활용할 수 있다고 생각하십니까?

문17 청소년 대상 빅데이터 체계 구축을 위해 선결되어야 하는 사항은 무엇이라고 생각하십니까?

문18 다양한 빅데이터 체계를 순차적으로 구성할 경우 우선적으로 먼저 구축해야 하는 주제는 무엇이라고 생각하십니까?(혹은 가장 중요하다고 생각하는 사항은 무엇인지요?)

문19 청소년 빅데이터 체계 구축을 위한 중장기 계획으로 어떠한 과정을 거치는 것이 바람직하다고 생각하십니까?

문20 청소년 빅데이터 체계 구축 및 활용의 발전방향은 무엇이라고 생각하십니까?

2. 전문가 설문조사지



청소년 빅데이터 체계 구축을 위한 전문가 설문조사지

바쁘신 중에 설문조사에 응해 주셔서 감사합니다. 본 설문 조사는 한국청소년정책연구원이 실시하며 여러분의 의견을 바탕으로 더 나은 청소년 정책·서비스 지원 체계를 마련하고자 이루어지고 있습니다. 응답내용은 연구목적으로만 활용될 것이며 익명으로 처리되기 때문에 비밀이 보장됩니다. 더 나은 서비스를 위해 끝까지 솔직하게 빠진 문항이 없도록 응답해 주세요.
감사합니다.

2020

한국청소년정책연구원 드림

☞ 들어가기에 앞서

빅데이터란? “빅데이터란 디지털 환경 등에서 생성되는 데이터로 그 규모가 방대하고, 생성 주기도 짧고, 형태도 수치 데이터뿐 아니라 문자와 영상 데이터를 포함하는 대규모 데이터를 말한다. 빅데이터 환경은 과거에 비해 데이터의 양이 폭증했다는 점과 함께 데이터의 종류도 다양해져 사람들의 행동은 물론 위치정보와 SNS를 통해 생각과 의견까지 분석하고 예측할 수 있다(네이버 지식백과). 행정안전부(2020)는 빅데이터의 특성을 4V로 설명하고 있는데 이는 다양성(Variety), 속도(Velocity), 대규모의 데이터 크기(Volume), 가치(Value)이다.”

※ 다음은 빅데이터 사용 경험과 욕구에 관련한 질문입니다. 자신에게 해당하는 란에 표시 또는 기재해 주세요.

1. 나는 중앙행정기관, 지방자치단체, 공공기관, 공공연구기관 등이 생성한 공공 영역의 빅데이터를 사용하여 연구 또는 분석을 실시한 경험이 있다.

_____ ① 예

_____ ② 아니오

1-1) 위에서 예라고 응답한 경우, 구체적으로 어떠한 공공 차원의 빅데이터를 사용하셨는지 있는 대로 적어주세요. 없는 경우 2번으로 넘어가 주세요.

예시) 서울특별시 서울시도시정책지표조사, 한국청소년정책연구원의 아동청소년인권실태조사, 과학기술정보통신부의 스마트폰 과의존 실태조사

1-2) 위에서 예라고 응답한 경우, 분석한 내용은 대략 어떠한 주제에 관한 것이었습니까?
기재하실 수 있는 범위 내에서 기재해주세요.

예시) 빈곤이 청소년 삶의 질에 미치는 영향, 게임사용이 청소년 건강에 미치는 영향

1-3) 위에서 예라고 응답한 경우, 공공 영역의 빅데이터를 사용하면서 불편하거나 개선
했으면 하는 점이 있으면 자유롭게 기재해 주세요.

2. 나는 포털이나 SNS, 인터넷 등 민간 영역의 빅데이터를 사용하여 연구 또는 분석을
실시한 경험이 있다.

_____ ① 예

_____ ② 아니오

2-1) 위에서 예라고 응답한 경우, 구체적으로 어떠한 민간 영역의 빅데이터를 사용하셨
는지 있는 대로 적어주세요. 없는 경우 3번으로 넘어가 주세요..

예시) 구글 트렌드 분석, 네이버 랩 분석, 트위터

2-2) 위에서 예라고 응답한 경우, 분석한 내용은 구체적으로 어떠한 주제에 관한 것이었
습니까? 기재하실 수 있는 범위 내에서 기재해주세요.

예시) 트위터 사용 여부에 따른 청소년 자존감 연구

2-3) 위에서 예라고 응답한 경우, 민간 영역의 빅데이터를 사용하면서 불편하거나 개선
했으면 했던 점이 있으면 자유롭게 기재해 주세요.

3. 귀하는 빅데이터에 대해 어느 정도 관심을 가지고 계십니까?

- _____ ① 전혀 관심 없다 _____ ② 별로 관심 없다 _____ ③ 보통
_____ ④ 관심 있다 _____ ⑤ 매우 관심 있다

4. 귀하는 앞으로 1년 이내에 빅데이터를 활용하여 연구 또는 분석을 할 생각이 있으십니까?

- _____ ① 그렇다 _____ ② 아니다

5. 4번에서 “그렇다”고 응답한 경우, 구체적으로 어떠한 데이터를 활용하여 어떠한 주제에 대해 연구하실 생각입니까?
(여러 개 있는 경우 있는 대로 적어 주세요)

5-1) 활용하고 싶은 데이터(구체적으로 기재해 주세요):

5-2) 연구하고 싶은 주제(구체적으로 기재해 주세요):

6. 4번에서 빅데이터를 활용하여 연구 또는 분석을 할 생각이 없다고 응답한 경우, 왜 빅데이터에 관심이 없거나 사용하실 생각이 없으신지 이유를 자유롭게 적어주세요.

7. 어떻게 하면 사람들이 빅데이터를 더 편리하고 쉽게 이용할 수 있다고 생각하시는지 의견이 있으시면 자유롭게 적어 주세요.

※ 다음은 청소년과 관련된 빅데이터 체계 구축을 위한 사항입니다. 자신의 의견에 가까운 란에 표시 또는 기재해 주세요.

8. 다음은 기존 연구(2011)에서 제시된 아동·청소년 핵심지표 영역체계 분류 영역(안)입니다. 청소년에 대한 빅데이터 체계를 구축하기 위한 분류 체계를 만든다고 가정할 경우 2020년 현재 적합하도록 아래 영역(안)에서 삭제, 추가 또는 수정·보완 되어야 한다고 생각하는 내용이 있으시면 자유롭게 적어주세요.

아동청소년 핵심지표 영역 체계 분류 영역(안)

1. 인구 및 가족, 2. 교육 및 역량, 3. 건강 및 심리, 4. 보호 및 안전,
5. 활동 및 문화, 6. 복지 및 권리, 7. 비행 및 범죄

9. 다음은 아동·청소년 핵심지표 영역체계 분류 영역(안)의 대분류 별 세부사항입니다. 청소년에 대한 빅데이터 체계를 구축하기 위한 분류 체계를 만든다고 가정할 경우 아래 세부사항에서 수정·보완 또는 추가 되어야 한다고 생각하는 사항이 있으시면 자유롭게 적어주세요.

9-1) 인구 및 가족 대분류

대분류	영역	세부영역	핵심지표
㉠ 인구 및 가족	1-1. 인구	1-1-1 아동·청소년 인구 규모	<ul style="list-style-type: none"> • 아동·청소년 인구 수 • 아동·청소년 인구구성비(전체 대비) • 아동·청소년 연령별(5세단위) 인구수 • 아동·청소년 성별, 지역별, 인구수(구성비)
		1-1-2 아동·청소년 인구 전망	<ul style="list-style-type: none"> • 아동·청소년 인구 추계
		1-1-3 아동·청소년 인구 동태	<ul style="list-style-type: none"> • 출생아 수 • 0세의 기대여명 • 청소년 출산율 • 저체중아 출산율 • 합계출산율 • 조혼인율 • 조이혼율 • 청소년 사망률 • 청소년 사망원인
	1-2. 가족형태	1-2-1 가족형태별 아동·청소년	<ul style="list-style-type: none"> • 양부모가정 아동·청소년수 • 한부모가정 아동·청소년수 • 재혼가정 아동·청소년수 • 조손가정 아동·청소년수 • 소년소녀가정 아동·청소년 수 • 입양가정 아동·청소년수
		1-2-2 다문화 아동·청소년	<ul style="list-style-type: none"> • 다문화가정 아동·청소년 수 • 다문화가정 자녀 재학 현황 • 탈북아동·청소년 수 • 탈북아동·청소년 재학 현황
	1-3. 가족관계	1-3-1 부모자녀관계	<ul style="list-style-type: none"> • 부모자녀간 동거여부 • 부모자녀간 의사소통 • 부모자녀간 함께 하는 활동
		1-3-2 부부관계	<ul style="list-style-type: none"> • 혼인상태(초혼/재혼/이혼) • 부부간 의사소통 • 부부갈등 빈도 및 강도 • 부부간 함께 하는 활동
		1-3-3 형제자매 관계	<ul style="list-style-type: none"> • 형제자매 수 • 형제자매간 의사소통 • 형제자매간 함께 하는 활동

수정·보완 또는 추가 되어야 한다고 생각하는 사항

9-2) 교육 및 역량 대분류

대분류	영역	세부영역	핵심지표
② 교육 및 역량	2-1 공교육	2-1-1 학생현황	<ul style="list-style-type: none"> • 학령인구수 • 진학률/취학률 • 학업중단률 • 대안학교 학생수 • 학교밖 청소년수
		2-1-2 학교현황	<ul style="list-style-type: none"> • 학교수/교원수 • 교원1인당 학생수 • 교육재정 규모 • 대안학교 학교수
		2-1-3 학업성취도	<ul style="list-style-type: none"> • 과목별 학업성적 • 국제학업성취도(PISA, TIMSS) • 자기주도적 학습시간
	2-2 사교육	2-2-1 사교육 규모	<ul style="list-style-type: none"> • 사교육 참여율 • 사교육비 • 학생1인당 사교육비 • 사교육시간
		2-2-2 사교육 효과	<ul style="list-style-type: none"> • 사교육효과성 • 사교육 만족도
	2-3 학교생활	2-3-1 학교생활	<ul style="list-style-type: none"> • 공부에 대한 흥미 • 학교 소속감 • 학교생활 만족도
		2-3-2 교사관계	<ul style="list-style-type: none"> • 교사-학생간 의사소통 • 교사와의 친밀도 • 교사만족도
		2-3-3 친구관계	<ul style="list-style-type: none"> • 친한 친구 수 • 친구와의 친밀도 • 친구와의 친밀도 • 동성/이성친구수
	2-4 핵심역량	2-4-1 지적 역량 (지적 도구 활용)	<ul style="list-style-type: none"> • 읽고, 쓰고 말하는 능력 • 수학적 능력 / 과학적 능력 • 정보활용능력 • 기술활용능력 • 사고력 / 창의력
		2-4-2 사회적 역량 (사회적 상호작용)	<ul style="list-style-type: none"> • 관계지향 • 협력, 협동 • 갈등관리
		2-4-3 자율적 역량 (자율적 행동)	<ul style="list-style-type: none"> • 자기관리 • 목표와 과제수립 • 자율성

수정·보완 또는 추가 되어야 한다고 생각하는 사항

9-3) 건강 및 심리 대분류

대분류	영역	세부영역	핵심지표
③ 건강 및 심리	3-1. 건강	3-1-1 신체건강	<ul style="list-style-type: none"> • 신장 및 체중 • 비만율 • 자체중율 • 체력(달리기, 멀리뛰기, 윗몸일으키기 등) • 신체건강상태(시력이상, 피부, 구강 질환 등) • 만성질환/장애여부 • 수면시간/습관 • 정기적인 운동습관 • 운동 횟수 및 시간
		3-1-2 영양섭취	<ul style="list-style-type: none"> • 영양소별 평균 섭취비율 • 식습관 • 결식율
		3-1-3 정신건강	<ul style="list-style-type: none"> • 스트레스 • 자살의도 • 우울증 • ADHD
		3-1-4 중독	<ul style="list-style-type: none"> • 인터넷 중독 • 게임 중독 • 휴대폰 중독 • 알콜 중독 • 약물 중독
	3-2. 심리	3-2-1 자기 이해	<ul style="list-style-type: none"> • 자아존중감 • 자기효능감 • 자기통제 • 자아정체감
		3-2-2 동기 및 정서	<ul style="list-style-type: none"> • 행복감 • 성취동기 • 정서장애(분노, 우울, 공격성, 불안) • 생활만족도

수정·보완 또는 추가 되어야 한다고 생각하는 사항

9-4) 보호 및 안전 대분류

대분류	영역	세부영역	핵심지표
4 보호 및 안전	4-1. 보호	4-1-1 유해업소, 약물 및 유해행위	<ul style="list-style-type: none"> • 유해업소 현황 • 유해업소 청소년 이용률 • 청소년 통행금지구역/제한구역 현황 • 유해업소 출입·고용 단속 현황 • 학교환경위생정화구역 현황 • 청소년 이용유해행위 단속 현황 • 유해약물 등 판매 단속 현황
		4-1-2 유해매체	<ul style="list-style-type: none"> • 영상물 심의현황 • 간행물 심의현황 • 음반 심의현황 • 청소년유해매체를 결정고시현황 • 청소년 스스로 지킴이(gp) 운영 현황 • 유해매체를 위반사범 단속현황
		4-1-3 성보호	<ul style="list-style-type: none"> • 청소년성매매 발생 및 검거 현황 • 성매매 피해자 발생 및 조치 현황 • 성매매 가해자 발생 및 조치 현황 • 신상정보 등록 및 열람대상자 현황 • 아동·청소년 대상 성범죄자 취업제한 현황
	4-2. 안전	4-2-1 학대경험	<ul style="list-style-type: none"> • 신체학대 유무 • 정서학대 유무 • 성학대 유무 • 방임 유무
		4-2-2 안전인식	<ul style="list-style-type: none"> • 안전인식 여부 • 위험물질 주의 정도 • 안전방비 착용 유무
		4-2-3 사고 발생	<ul style="list-style-type: none"> • 교통사고 발생률 • 안전사고 발생률

수정·보완 또는 추가 되어야 한다고 생각하는 사항

9-5) 활동 및 문화 대분류

대분류	영역	세부영역	핵심지표
⑤ 활동 및 문화	5-1. 활동	5-1-1 청소년단체활동	<ul style="list-style-type: none"> • 청소년단체 회원 가입 여부 • 청소년단체활동 참여율 • 청소년단체활동 만족도
		5-1-2 동아리활동	<ul style="list-style-type: none"> • 동아리 회원 가입 여부 • 동아리활동 참여율 • 동아리활동 만족도
		5-1-3 자치활동	<ul style="list-style-type: none"> • 자치활동 임원 참여 여부 • 자치활동 참여율 • 자치활동 만족도
		5-1-4 자원봉사활동	<ul style="list-style-type: none"> • 자원봉사활동 참여율 • 자원봉사활동 시간 • 자원봉사활동 만족도
		5-1-5 청소년수련시설 활동	<ul style="list-style-type: none"> • 청소년수련시설 이용률 • 단체수련활동(캠프) 참여율 • 단체수련활동(캠프) 만족도
		5-1-6 국제교류활동 국제교류활동참여율	<ul style="list-style-type: none"> • 국제교류활동참여율
		5-1-7 방과후활동	<ul style="list-style-type: none"> • 방과후 나홀로 아동수 • 방과후 주도 하는 일 • 방과후 함께있는 사람 • 방과후 학교 이용경험/만족도 • 지역아동센터 이용경험/만족도 • 방과후아카데미 이용경험/만족도
	5-2. 문화	5-2-1 견학	<ul style="list-style-type: none"> • 박물관/미술관/문화재 이용률
		5-2-2 공연/스포츠관람	<ul style="list-style-type: none"> • 영화/연극/뮤지컬/콘서트 관람률 • 스포츠경기 관람률
		5-2-3 매체	<ul style="list-style-type: none"> • TV/비디오/라디오 시청시간 • 신문/잡지 구독시간 • 인터넷/컴퓨터 이용시간 • 휴대폰/MP3 플레이어 이용시간 • 컴퓨터 게임시간
		5-2-4 독서/도서관	<ul style="list-style-type: none"> • 도서관 이용률 • 독서시간 • 독서율 • 평균독서권수
		5-2-5 여가생활	<ul style="list-style-type: none"> • 여가시간 • 여가생활만족도 • 여가활동 순위

수정·보완 또는 추가 되어야 한다고 생각하는 사항

9-6) 복지 및 보호 대분류

대분류	영역	세부영역	핵심지표
⑥ 복지 및 권리	6-1. 복지	6-1-1 취약, 위기청소년	<ul style="list-style-type: none"> • 아동빈곤율(절대, 상대빈곤율) • 1달러 미만 생활 아동비율 • 국민기초생활수급 아동 비율 • 취약계층 아동청소년 규모 • 요보호 아동청소년 규모 • 위기 아동청소년 규모 • 가출청소년 규모 • 학업중단 청소년 규모
		6-1-2 복지시설 이용	<ul style="list-style-type: none"> • 아동·청소년복지시설 이용 경험/만족도 • 아동·청소년보호시설 이용 경험/만족도 • 청소년쉼터 이용 경험/만족도
		6-1-3 복지서비스 이용	<ul style="list-style-type: none"> • 상담서비스 이용/만족도 • 의식주 서비스 이용/만족도
		6-1-4 복지인식	<ul style="list-style-type: none"> • 복지시설에 대한 인지도 • 복지서비스에 대한 인지도
	6-2. 권리	6-2-1 권리 및 참여 의식	<ul style="list-style-type: none"> • 아동 권리에 대한 인지도 • 권리 주장 방법 • 인권에 대한 만족도
		6-2-2 학생권리	<ul style="list-style-type: none"> • 학교 규칙, 절차 참여도 • 학교의 인권보호 여부 • 학교활동 선택 정도 • 수업 선택정도
		6-2-3 인권인식	<ul style="list-style-type: none"> • 신체에 대한 인권 인식정도 • 집회/결사 인권 인식정도 • 자치활동 인권 인식정도
		6-2-4 인권교육	<ul style="list-style-type: none"> • 인권교육 경험여부 • 인권교육에 대한 만족도
		6-2-5 처벌, 징계처리	<ul style="list-style-type: none"> • 학교규칙 처벌 방법 • 처벌, 징계의 적합성 여부

수정·보완 또는 추가 되어야 한다고 생각하는 사항

9-7) 비행 및 범죄 대분류

대분류	영역	세부영역	핵심지표
7 비행 및 범죄	7-1 비행	7-1-1. 음주/흡연/약물	<ul style="list-style-type: none"> • 흡연 경험율 • 음주 경험율 • 본드 등 흡입제 사용 여부 • 유해약물 경험율
		7-1-2 학교폭력	<ul style="list-style-type: none"> • 학교폭력 피해경험 여부 • 집단따돌림 피해경험 여부 • 가해경험, 가해종류, 인식 • 가해후 처벌, 상담, 치료 여부
		7-1-3 가출	<ul style="list-style-type: none"> • 가출청소년 현황 • 가출 후 경험행동 • 가출에 대한 인식 • 가출후 연계시설, 상담 이용여부
		7-1-4. 성관련 경험	<ul style="list-style-type: none"> • 성교육 유무, 시기, 도움정도 • 성 관련 의논상대 • 성에 대한 태도 • 성관계 경험 유무 • 임신 및 낙태 경험여부
	7-2 범죄	7-2-1. 소년범죄 경험	<ul style="list-style-type: none"> • 청소년 범죄 현황 • 학생 범죄 현황 • 소년범죄/학생범죄 처리 현황 • 기소유예 처분 청소년 현황 • 청소년 보호관찰 대상 현황 • 청소년보호관찰대상자 추이 • 소년원 현황 • 소년분류심사원 현황

수정·보완 또는 추가 되어야 한다고 생각하는 사항

10. 추후 빅데이터를 통해 청소년 정책, 서비스의 미래방향, 청사진을 예측한다고 가정할 경우 우선적으로 어떠한 데이터가 구축되어야 한다고 생각하십니까?(데이터의 유형, 내용 등) 의견을 자유롭게 적어 주세요.

11. 아동·청소년 분야에서 빅데이터를 보다 효과적으로 활용하기 위해 선행되어야 할 일은 무엇이라고 생각하십니까?

12. 이 외에 아동·청소년 분야에서 빅데이터 체계가 효과적으로 구축될 수 있도록 하기 위한 아이디어가 있으시면 자유롭게 기재해 주세요.

※ 다음은 귀하에 대한 질문입니다.

13. 성별은? _____ ① 남자 _____ ② 여자

14. 연령은? 만으로 _____ 세

15. 다음 중 어디에 해당합니까?

_____ ① 현장전문가 _____ ② 공무원
_____ ③ 교수, 연구원 _____ ④ 기타(구체적으로 적어주세요_____)

16. 아동·청소년 또는 빅데이터 관련 관학연·현장 분야에서 일하신 총 경력은 약 몇 년입니까?

약 _____ 년

17. 학력은?

_____ ① 고졸 _____ ② 대학재학/졸업 _____ ③ 석사재학/졸업
_____ ④ 박사재학/졸업 _____ ⑤ 기타(구체적으로 적어주세요_____)

♣ 설문이 모두 끝났습니다. 긴 시간 동안 설문에 응답해 주셔서 감사합니다. ♣

3. 아동청소년관련 승인통계 변인 조사표(예)82)

1) 중앙행정기관

번호	작성기관	대분류	중분류	소분류
1	검찰청	범죄분석통계	가정·학교폭력 처분결과	가정폭력 처분결과
2	검찰청	범죄분석통계	가정·학교폭력 처분결과	학교폭력 처분결과
3	검찰청	범죄분석통계	공무원범죄분석	공무원범죄자 처분결과
4	검찰청	범죄분석통계	공무원범죄분석	공무원범죄자 처분결과 (법법별, 처분결과별)(분기)
5	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(총괄)
6	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(서울중앙지방검찰청)
7	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(서울남부지방검찰청)
8	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(서울동부지방검찰청)
9	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(서울북부지방검찰청)
10	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(서울서부지방검찰청)
11	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(부산지방검찰청)
12	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(대구지방검찰청)
13	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(인천지방검찰청)
14	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(광주지방검찰청)
15	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(대전지방검찰청)
16	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(울산지방검찰청)
17	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(수원지방검찰청)
18	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(의정부지방검찰청)
19	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(춘천지방검찰청)
20	검찰청	범죄분석통계	범죄자처분결과	범죄자 처분결과(청주지방검찰청)

82) 분량 관계상 기관유형별 예시 1장씩만 제시

2) 지방자치단체

번호	작성기관	대분류	중분류	소분류
1	서울특별시	서울시도시정책지표조사	가구주	가사노동 분담률
2	서울특별시	서울시도시정책지표조사	가구주	가족안전유형
3	서울특별시	서울시도시정책지표조사	가구주	가족안전유형
4	서울특별시	서울시도시정책지표조사	가구주	가족 안전 위험 여부
5	서울특별시	서울시도시정책지표조사	가구주	가족 안전 위험 여부
6	서울특별시	서울시도시정책지표조사	가구주	과거 5년동안 이사경험 및 변화
7	서울특별시	서울시도시정책지표조사	가구주	교육환경 만족도
8	서울특별시	서울시도시정책지표조사	가구주	교육환경 만족도
9	서울특별시	서울시도시정책지표조사	가구주	교육환경 만족도
10	서울특별시	서울시도시정책지표조사	가구주	부동산 자산 비율
11	서울특별시	서울시도시정책지표조사	가구주	부채여부
12	서울특별시	서울시도시정책지표조사	가구주	부채여부
13	서울특별시	서울시도시정책지표조사	가구주	부채이유
14	서울특별시	서울시도시정책지표조사	가구주	부채이유
15	서울특별시	서울시도시정책지표조사	가구주	부채이유
16	서울특별시	서울시도시정책지표조사	가구주	생활환경 만족도
17	서울특별시	서울시도시정책지표조사	가구주	생활환경 만족도
18	서울특별시	서울시도시정책지표조사	가구주	거주지 선택 시 고려요인
19	서울특별시	서울시도시정책지표조사	가구주	주거지 선택 고려요인
20	서울특별시	서울시도시정책지표조사	가구주	주거지 선택 고려요인

3) 공사, 공단

번호	작성기관	대분류	중분류	소분류
1	국민건강보험공단	건강검진통계	일반건강검진	연령별 성별 일반건강검진 1차 판정현황(전체)
2	국민건강보험공단	건강검진통계	일반건강검진	연령별 성별 일반건강검진 1차 판정현황(외국인)
3	국민건강보험공단	건강검진통계	일반건강검진	연령별 성별 일반건강검진 1차 판정현황(전체)
4	국민건강보험공단	건강검진통계	일반건강검진	연령별 성별 일반건강검진 1차 판정현황(외국인)
5	국민건강보험공단	건강검진통계	일반건강검진	연령별 성별 일반건강검진 1차 판정현황
6	국민건강보험공단	건강검진통계	일반건강검진	연령별 성별 일반건강검진 1차 판정결과(정상B)
7	국민건강보험공단	건강검진통계	일반건강검진	연령별 성별 일반건강검진 1차 판정결과(질환의심)
8	국민건강보험공단	건강검진통계	일반건강검진	연령별 성별 일반건강검진 1차 판정결과(유질환자)
9	국민건강보험공단	건강검진통계	일반건강검진	시도별 연령별 성별 일반건강검진 정신건강검사 결과 (기분상태 및 우울증)
10	국민건강보험공단	건강검진통계	일반건강검진	연령별 성별 일반건강검진 1차 판정결과 세부현황
11	국민건강보험공단	건강검진통계	일반건강검진	연령별 성별 일반건강검진 2차 판정현황
12	국민건강보험공단	건강검진통계	일반건강검진	연령별 성별 일반건강검진 2차 판정현황(당뇨병, 고혈압)
13	국민건강보험공단	건강검진통계	암검진	연령별 성별 암검진 대상 및 수검인원 현황

번호	작성기관	대분류	중분류	소분류
14	국민건강보험공단	건강검진통계	암검진	연령별 자궁경부암 검진 결과 판정 현황
15	국민건강보험공단	건강검진통계	구강검진	연령별 성별 구강검진 대상 및 수검인원 현황
16	국민건강보험공단	건강검진통계	구강검진	연령별 성별 일반건강검진 대상자 구강검진 종합소견 현황
17	국민건강보험공단	건강검진통계	문진	연령별 성별 검진대상 질환력 (과거력, 가족력) 문진결과
18	국민건강보험공단	건강검진통계	문진	연령별 성별 검진대상 흡연 관련 문진결과
19	국민건강보험공단	건강검진통계	문진	연령별 성별 전자흡연 관련 문항
20	국민건강보험공단	건강검진통계	문진	연령별 성별 음주 관련 문항

4) 연구기관

번호	작성기관	대분류	소분류
1	한국교육개발원	교육기관(고등교육기관및직업계고) 졸업자취업통계(2011년 이후)	하반기 학제별 설립별 졸업상황
2	한국교육개발원	교육기관(고등교육기관및직업계고) 졸업자취업통계(2011년 이후)	하반기 계열별 졸업상황
3	한국교육개발원	교육기관(고등교육기관및직업계고) 졸업자취업통계(2011년 이후)	하반기 진학상황
4	한국교육개발원	교육기관(고등교육기관및직업계고) 졸업자취업통계(2011년 이후)	졸업상황 및 건강보험 DB연계 취업상황
5	한국교육개발원	교육기관(고등교육기관및직업계고) 졸업자취업통계(2011년 이후)	계열별 졸업상황 및 건강보 험DB연계 취업상황
6	한국교육개발원	교육기관(고등교육기관및직업계고) 졸업자취업통계(2011년 이후)	진학상황
7	한국교육개발원	교육기관(고등교육기관및직업계고) 졸업자취업통계(2011년 이후)	졸업상황
8	한국교육개발원	교육기관(고등교육기관및직업계고) 졸업자취업통계(2011년 이후)	진학상황
9	한국교육개발원	교육기관(고등교육기관및직업계고) 졸업자취업통계(2011년 이후)	계열별 취업상황
10	한국교육개발원	교육기관(고등교육기관및직업계고) 졸업자취업통계(2011년 이후)	취업상황
11	한국교육개발원	교육기관(고등교육기관및직업계고) 졸업자취업통계(2011년 이후)	미취업상황
12	한국교육개발원	교육기관(고등교육기관및직업계고) 졸업자취업통계(2011년 이후)	학교급별 취업현황
13	한국교육개발원	교육기본통계	유치원 개황
14	한국교육개발원	교육기본통계	초등학교 개황
15	한국교육개발원	교육기본통계	중학교 개황
16	한국교육개발원	교육기본통계	고등학교 개황
17	한국교육개발원	교육기본통계	특수학교 개황
18	한국교육개발원	교육기본통계	고등공민학교 개황
19	한국교육개발원	교육기본통계	고등기술학교 개황
20	한국교육개발원	교육기본통계	각종학교 개황

ABSTRACT

This study investigates the possibility of establishing a big data system for adolescents, derives suggestions for establishing a system based on the results, and then presents basic data for the establishment of a big data system for adolescents. It also seeks to identify how it is possible to predict important youth-related issues through analysis using structured big data and unstructured social big data. The study then proposes specific measures on how a big data system for adolescents can be used to predict youth-related changes and policy demands in the future, detect risk factors in early stages, and proactively prevent further danger. To this end, a domestic and foreign literature review, big data analytics colloquium, foreign big data system establishment and utilization case studies, youth-centered ontology development, professional advice on big data system establishment and utilization, expert in-depth interviews, expert surveys, and policy conference were all used as research methods. The major findings of this study are as follows. First, in the in-depth interviews with experts, respondents placed an emphasis on life design and life modification perspectives with respect to the issues that the big data for adolescents should deal with. Second, the expert survey results showed that the restriction of access due to the protection of personal information accounted for the highest proportion among matters which are inconvenient and need to be improved when using big data in the public sphere, at 37.5%. Based on the research results, this study proposed methods to establish a big data system for adolescents and its subsequent utilization measures.

2020년 한국청소년정책연구원 발간자료 목록

기관고유과제

- 20-R01 청소년의 혐오표현 노출실태 및 대응 방안 연구 / 김영한·이유진·조아미·임성택
- 20-R02 청소년·청년 디지털 플랫폼노동 실태 및 대응방안 연구 / 유민상·최정원·이수정·장혜림
- 20-R03 Z세대 10대 청소년의 가치관 변화 연구 / 오해섭·문호영
- 20-R04 청소년활동의 사회적 가치 제고방안 연구 / 최용환·성유리·박윤수·김보경
- 20-R05 청소년지도자 양성시스템 재구축방안 연구 II : 청소년상담사를 중심으로 / 최창욱·좌동훈·이종원·남화성·정지희
- 20-R06 청년 사회·경제실태 및 정책방안 연구V / 김형주·연보라·배정희
- 20-R06-1 청년 사회·경제실태 및 정책방안 연구V - 전년도 공개데이터 심층분석연구 / 김형주·연보라·배정희
- 20-R06-2 청년 사회·경제실태 및 정책방안 연구V - 기초분석보고서 / 김형주·연보라·배정희
- 20-R07 국가 미래인적자원으로서 재외동포청소년 성장과 지원방안 연구 III / 김경준·김정숙·윤철경
- 20-R08 청소년 빅데이터 체계 구축 및 활용방안 연구 / 서정아·성윤숙·송태민
- 20-R08-1 청소년 빅데이터 체계 구축 및 활용방안 연구 데이터 분석 보고서 1_청소년 비만에 대한 비정형 빅데이터 연구 / 송태민
- 20-R08-2 청소년 빅데이터 체계 구축 및 활용방안 연구 데이터 분석 보고서 2_청소년 행복 결정요인에 대한 정형 빅데이터 연구 / 홍성호
- 20-R09 청소년의 주거권 실태와 보장방안 연구: 사회배제 관점을 중심으로 / 김지연·김승경·임세희·최은영
- 20-R10 2020 아동·청소년 권리에 관한 국제협약 이행 연구 - 한국 아동·청소년 인권실태 : 총괄보고서 / 김영지·황세영·최홍일·이민희·김진호
- 20-R10-1 2020 아동·청소년 권리에 관한 국제협약 이행 연구 - 한국 아동·청소년 인권실태 : 심화분석보고서 / 박환보·주경필
- 20-R10-2 2020 아동·청소년 권리에 관한 국제협약 이행 연구 - 한국 아동·청소년 인권실태 : 기초분석보고서 / 김영지·황세영·최홍일
- 20-R11 지역사회 네트워크를 활용한 청소년 성장지원 정책 추진체계 구축 방안 연구 II / 최인재·강경균·송민경·조윤정·김가희
- 20-R12 2020 한국아동·청소년패널조사 : 사업보고서 / 하형석·황진구·김성은·이용해
- 20-R12-1 2020 한국아동·청소년패널조사 : 데이터분석보고서 / 김성은·황영식

- 20-R13 2020 다문화청소년 종단연구 : 총괄보고서 / 양계민·장윤선·정윤미
 20-R13-1 2020 다문화청소년 종단연구 : 기초분석보고서: 1기패널 / 양계민·장윤선·정윤미
 20-R13-2 2020 다문화청소년 종단연구 : 기초분석보고서: 2기패널 / 양계민·장윤선·정윤미

협동연구과제

- 경제·인문사회연구회 협동연구총서 20-82-01 학교 밖 청소년 지역사회 지원방안 연구 III :
 질적패널 조사를 중심으로 / 김희진·장근영·이동훈·윤철경 (자체번호 20-R14)
 경제·인문사회연구회 협동연구총서 20-82-02 학교 밖 청소년 지역사회 지원방안 연구 III : 학교
 밖 여성 청소년을 중심으로 / 오은진·장희영 (자체번호 20-R14-1)
 경제·인문사회연구회 협동연구총서 20-83-01 청년 핵심정책 대상별 실태 및 지원방안 연구 III :
 청년 이직자 - 총괄보고서 / 김기현·신동훈·변금선·고혜진·신인철 (자체번호 20-R15)
 경제·인문사회연구회 협동연구총서 20-83-02 청년 핵심정책 대상별 실태 및 지원방안 연구 III :
 청년 이직자 - 심층분석보고서 / 김기현·신동훈·고혜진·신인철 (자체번호 20-R15-1)
 경제·인문사회연구회 협동연구총서 20-84-01 청소년의 건강권 보장을 위한 정책방안 연구 II :
 위기청소년 / 백해정·임희진·송미경·김양희 (자체번호 20-R16)
 경제·인문사회연구회 협동연구총서 20-84-02 청소년의 건강권 보장을 위한 정책방안 연구 II :
 위기청소년-국내·외 주요 법령 및 건강정책 분석 / 류정희·이상정·박선영·전민경 (자체번호
 20-R16-1)
 경제·인문사회연구회 협동연구총서 20-84-03 청소년의 건강권 보장을 위한 정책방안 연구 II :
 위기청소년 - 기초분석보고서 / 백해정·임희진 (자체번호 20-R16-2)
 경제·인문사회연구회 협동연구총서 20-85-01 청소년 미디어 이용 실태 및 대상별 정책대응방안
 연구 I : 초등학교 / 배상률·이창호·이정림 (자체번호 20-R17)
 경제·인문사회연구회 협동연구총서 20-85-02 청소년 미디어 이용 실태 및 대상별 정책대응방안 연구 I :
 초등학교 - 해외사례 조사 / 정현선·심우민·윤지원·김광희·최원석 (자체번호 20-R17-1)
 경제·인문사회연구회 협동연구총서 20-85-03 청소년 미디어 이용 실태 및 대상별 정책대응방안
 연구 I : 초등학교 - 기초분석보고서 / 배상률·이창호 (자체번호 20-R17-2)

연구개발적립금

- 20-R24 위기청소년 현황 및 실태조사 기초연구 / 황여정·이정민
 20-R24-1 위기청소년 현황 및 실태조사 기초연구: 예비조사 데이터분석보고서 / 황여정·이정민·김수혜
 20-R25 10대 청소년 포럼 운영 / 모상현
 20-R26 코로나 19 확산 및 이후 사회변화에 따른 청소년정책의 대응방안 / 김현철

수 시 과 제

- 20-R18 학생 인권교육을 위한 현장 실천 강화 방안 연구 / 이정민·이종원
- 20-R19 청년정책 평가체계 구축을 위한 기초연구 / 배정희·김기현
- 20-R20 가출청소년 지원 강화를 위한 청소년복지시설 재구조화 연구 / 황진구·김지연
- 20-R21 선거법 개정에 따른 청소년 정책 및 활동 지원 방안 연구 / 이창호
- 20-R22 청소년지도사 자격검정과목 개편 방안 연구 / 김경준·이종원·박정배
- 20-R23 학교밖청소년지원센터 운영모형 개발 / 김희진·백혜정

수 탁 과 제

- 20-R27 청소년 비즈클 활성화를 위한 고교교육 정책 연계방안 연구 / 강경균·안재영·황은희
- 20-R28 청년정책 현황 진단 및 정책추진 실효성 제고 방안 연구 / 김기현·유민상·변금선·배정희
- 20-R29 (가칭)청양군 청소년재단 설립타당성 연구 / 김영한
- 20-R29-1 (가칭)청양군 청소년재단 설립타당성 연구(요약본) / 김영한
- 20-R30 2020년 학교폭력 예방교육 컨설팅 매뉴얼 / 성윤숙·양미석
- 20-R31 학교폭력 예방 어울림 기본 프로그램(5종)
- 20-R31-1 학교폭력 예방 어울림 기본 프로그램(초등학교 1~2학년용) / 성윤숙·서정아·장윤선·서고은·김성은
- 20-R31-2 학교폭력 예방 어울림 기본 프로그램(초등학교 3~4학년용) / 성윤숙·서정아·장윤선·서고은·김성은
- 20-R31-3 학교폭력 예방 어울림 기본 프로그램(초등학교 5~6학년용) / 성윤숙·서정아·장윤선·서고은·김성은
- 20-R31-4 학교폭력 예방 어울림 기본 프로그램(중학교용) / 성윤숙·서정아·장윤선·서고은·김성은
- 20-R31-5 학교폭력 예방 어울림 기본 프로그램(고등학교용) / 성윤숙·서정아·장윤선·서고은·김성은
- 20-R32 2019 학교폭력 예방 어울림 프로그램 적용효과 분석 / 성윤숙·양미석
- 20-R33 지역단위 학교폭력 예방교육 프로그램 운영현황 및 적용방안: 회복적 생활교육을 중심으로 / 성윤숙·양미석
- 20-R34 2018~2019 학교폭력 예방교육 연구학교 결과보고서
- 20-R34-1 2018~2019 학교폭력 예방교육 연구학교 결과보고서(초등학교) / 성윤숙·이윤소
- 20-R34-2 2018~2019 학교폭력 예방교육 연구학교 결과보고서(중학교) / 성윤숙·이윤소
- 20-R34-3 2018~2019 학교폭력 예방교육 연구학교 결과보고서(고등학교) / 성윤숙·이윤소
- 20-R35 교육과정 기반 어울림 프로그램 운영 사례 / 성윤숙·이윤소

- 20-R36 어울림 자유학기 프로그램(5종)
- 20-R36-1 어울림 자유학기 주제선택 프로그램(놀러와! 어울림 세상) /
성윤숙·이선희·정미선·선보라·이혜옥·이윤소
- 20-R36-2 자유학기 프로그램 동아리활동 / 성윤숙·배은정·이윤소
- 20-R36-3 어울림 자유학기 교과연계 주제선택 프로그램(사회) / 성윤숙·선보라·이윤소
- 20-R36-4 어울림 자유학기 교과연계 주제선택 프로그램(도덕) / 성윤숙·이혜옥·이윤소
- 20-R36-5 어울림 자유학기 교과연계 주제선택 프로그램(국어) / 성윤숙·이선희·정미선·이윤소
- 20-R37 외국의 학교폭력 예방교육 정책 및 프로그램 동향 / 성윤숙
- 20-R38 서울특별시 청소년시설 종합성과평가 평가지표 개선 용역 / 김형주·김혁진·김정주
- 20-R39 제1차 청년정책 기본계획 수립 연구 / 김기현·유민상·변금선·배정희
- 20-R40 학교 안·밖 청소년정책 협력체계 구성 및 연계방안 연구 / 최창욱·좌동훈·남화성
- 20-R41 소년원학생 재범방지 통합프로그램 개발 연구 / 이유진
- 20-R41-1 꿈과 친구사이 : 소년원학생 재범방지 통합프로그램 매뉴얼 / 이유진
- 20-R42 청소년 주도적인 활동지원을 위한 법률개정 방안 연구 / 최창욱·문호영
- 20-R43 보호종료아동 주거지원 통합서비스 사업 사례관리사 및 자립업무종사자 직무분석 연구 / 김지연·백혜정·이상정
- 20-R44 2020년도 청소년 인터넷 건전이용제도 적용 게임물 평가 / 배상률·유홍식·김동일
- 20-R45 청소년 교육·문화공간 확충을 위한 지자체-교육청 협력 강화 방안 /
김영지·황세영·손진희·박명선·박종원·조기영
- 20-R46 2020년 청소년특별회의 정책과제 연구 / 최창욱·좌동훈
- 20-R47 농업인력 유입 확대를 위한 농촌 청소년 실태조사 / 오해섭·최홍일
- 20-R48 성남시 청소년시설 확충 및 개선 연구 / 황진구·남화성
- 20-R49 다함께돌봄센터 현장적용 프로그램 개발연구 / 황진구·좌동훈
- 20-R50 2020년 청소년수련시설종합평가 / 김경준·김영지·최창욱
- 20-R51 지역센터 관리운영을 위한 평가지표 개선방안 / 최용환·곽창규·이성규
- 20-R52 수원시 위기청소년 실태조사 연구 / 최용환·김보경
- 20-R53 2020 청소년방과후아카데미 효과·만족도 연구 / 양계민
- 20-R54 2020년 학교 내 대안교실 모니터링 연구 / 최인재·송원일·박지원
- 20-R55 2020년 대안교육 위탁교육기관 실태조사 연구 / 최인재·송원일·배수인
- 20-R56 복권기금 꿈사다리 장학생의 성장관리 및 성과측정 방안 연구 / 이정민·성유리·김혜원
- 20-R57 2020년 청소년 매체이용 및 유해환경 실태조사 / 김지연·김승경·백혜정·황여정
- 20-R57-1 2020년 청소년 매체이용 및 유해환경 실태조사(위기청소년 결과 분석) /
김지연·김승경·백혜정·황여정
- 20-R57-2 2020년 청소년 매체이용 및 유해환경 실태조사(부록: 기초통계결과표) /
김지연·김승경·백혜정·황여정

- 20-R58 지역사회 협력을 통한 직업계고 혁신 지원 방안 연구 - 특성화고 혁신지원 운영모형 개발 / 강경균·김영만·김용성
- 20-R59 지역아동센터 아동패널조사 2020 / 김희진·임희진·정윤미
- 20-R60 이주배경 아동·청소년 지원 지역기관 연계 시범사업 모니터링, 평가 및 모델 개발 / 연보라·최정원·김성은
- 20-R61 학업중단 현황 및 지원 방안 / 김성은·박하나·김현수
- 20-R62 환경 변화에 따른 후기청소년 정책 발전 방향 연구 / 장근영·김기헌

세미나 및 워크숍 자료집

- 20-S01 2020년 학교폭력 예방교육 컨설팅단 워크숍(1.30)
- 20-S02 2019년 학교폭력 예방교육 운영학교 성과보고회(1.31)
- 20-S03 2020년 제28회 청소년정책포럼 : 학교밖 청소년 취업 및 자립지원 방안(1.22)
- 20-S04 2019년 학교폭력 예방교육 어울림 프로그램 운영 우수사례집(1.31)
- 20-S05 제25차 NYPI 직원 역량강화 콜로키움: 강원국 교수의 매력적인 글쓰기 특강(4.23)
- 20-S06 제26차 NYPI 직원 역량강화 콜로키움: 김병완 작가의 쿼터 독서법 및 책쓰기 특강(4.21)
- 20-S07 2020년 제29회 청소년정책포럼 : 학교 밖 청소년 규모 추정 방안(6.19)
- 20-S08 2020년 제30회 청소년정책포럼 : 18세 선거권 이후 청소년 정치교육의 방향 및 과제(7.17)
- 20-S09 2020년 학교폭력 예방교육 컨설팅단 실무 역량 강화 연수(초등, 중등, 고등)(7.28)
- 20-S10 2020년 제31회 청소년정책포럼 : 청소년정책 관련 법제도 개선 방향(7.28)
- 20-S11 제30차 NYPI 직원 역량강화 콜로키움: 플랫폼 노동의 미래를 묻다 <별점 인생>(8.5)
- 20-S12 2020년 학교폭력 예방교육 컨설팅단 상반기 성과보고회(8.11)
- 20-S13 제31차 NYPI 직원 역량강화 콜로키움: Z세대의 SNS 마켓에서의 소비자사회화 경험에 대한 현상학적 연구: 인스타마켓을 중심으로(8.26)
- 20-S14 2020년 제32회 청소년정책포럼 : 포스트 코로나 시대 청소년 활동과 정책의 방향(8.18)
- 20-S15 청소년 이슈 관련 네트워크 분석(8.25)
- 20-S16 2020년 제33회 청소년정책포럼 : 학교 안밖 청소년 협력체계 구성 연계 방안(9.23)
- 20-S17 제32차 NYPI 직원 역량강화 콜로키움: 청소년의 혐오표현 노출실태 및 대응방안(10.6)
- 20-S18 2020년 대안학교 관리자 협의회 및 담당자 역량 강화 워크숍 (10.5)
- 20-S19 2020년 학교 내 대안교실 모니터링 위원 워크숍(10.5)
- 20-S20 2020년 제34회 청소년정책포럼 : 코로나 시대 청소년 성장지원의 방향(10.21)
- 20-S21 2020 유료필로조피 서울 대회(11.12~13)
- 20-S22 2020년 미인가 대안교육시설 관리자 및 담당자 역량 강화 온라인 워크숍(11.13)
- 20-S23 제35차 청소년정책포럼: Z세대 청소년의 가치관 변화와 정책적 대응방안(11.18)

- 20-S24 2020년 학교폭력 예방교육 발전방안 포럼 (해외 사이버폭력 정책 동향 및 국내 사이버폭력 예방교육 방향 모색)(12.18)
- 20-S25 제3회 한일진로교육포럼(11.28)
- 20-S26 제9회 한국아동·청소년패널 학술대회 자료집(11.20)
- 20-S27 제36차 청소년정책포럼: 우리나라 한국아동·청소년의 삶은 어떠한가?(12.1)
- 20-S28 제37차 청소년정책포럼: 이주배경 아동·청소년의 지원 체계의 과제와 방향(12.8)
- 20-S29 제1차 10대청소년목소리포럼: 코로나로 세겨진 우울의 시대 우리의 꿈과 이상을 쓰다 -동고동락, 각양각색 청소년의 희망이야기(12.5)
- 20-S30 제38차 청소년정책포럼: 후기청소년 지원정책의 과제와 방향(12.17)
- 20-S31 제39차 청소년정책포럼: 청소년 매체이용 및 유해환경 실태조사 주요 결과와 과제(12.22)
- 20-S32 진로위기학생 유형별 진로교육 실천과제 성과보고회 자료집(11.27)

학 술 지

- 「한국청소년연구」 제31권 제1호(통권 제96호)
- 「한국청소년연구」 제31권 제2호(통권 제97호)
- 「한국청소년연구」 제31권 제3호(통권 제98호)
- 「한국청소년연구」 제31권 제4호(통권 제99호)

기타 발간물

- NYPI Bluenote 이슈 & 정책 119호 : 대안교육 위탁교육기관의 질 제고 방안
- NYPI Bluenote 이슈 & 정책 120호 : 청년 핵심정책 대상별 실태 및 지원방안 연구 II : 학교 졸업예정자
- NYPI Bluenote 이슈 & 정책 121호 : 민간 위탁형 공립 대안학교의 운영 현황 및 발전 과제
- NYPI Bluenote 이슈 & 정책 122호 : 국가 미래인적자원으로서 재외동포청소년 성장과 지원방안 연구
- NYPI Bluenote 이슈 & 정책 123호 : 다문화청소년의 심리·사회적 지원방안 모색
- NYPI Bluenote 이슈 & 정책 124호 : 정보기술을 활용한 위기청소년 사회서비스 확충 방안
- NYPI Bluenote 이슈 & 정책 125호 : 미래지향적 청소년관계법 정비 방안
- NYPI Bluenote 이슈 & 정책 126호 : 「청년기본법」제정 전·후 정책형성의 성과와 한계, 그리고 앞으로의 과제
- NYPI Bluenote 이슈 & 정책 127호 : 청년예술인 지위 및 권리보장을 위한 '예술 활동 증명' 기준의 적정성 검토
- NYPI Bluenote 이슈 & 정책 128호 : 청년종합실태조사(안) 시행을 위한 조사 설계 방향

NYPI Bluenote 이슈 & 정책 129호 : 정부 및 지자체 청년 정책참여 현황과 과제

NYPI Bluenote 통계 49호 : 청소년 '일 경험' 제도 운영 실태 및 정책방안 연구 II : 대학교 실습학기제를
중심으로

NYPI Bluenote 통계 50호 : '학교 밖 청소년'이 얼마나 있을까?

NYPI Bluenote 통계 51호 : 청소년지도사 양성제도 관련 실태조사

NYPI Bluenote 통계 52호 : 한국아동·청소년패널조사 2018 제2차년도 주요 조사 결과

연구에 도움을 주신 분들

◆ 집 필 진 ◆

- 류 시 원 (인제대학교 보건대학원 교수)
손 병 덕 (충신대학교 사회복지학과 교수)
천 정 응 (대구가톨릭대학교 사회복지학과 교수)

◆ 자 문 · 협 력 진 ◆

- 김 창 환 (한국교육개발원 선임연구위원)
오 미 애 (한국보건사회연구원 빅데이터·정보연구센터장)
이 민 희 (평택대학교 아동청소년교육상담학과 교수)
이 희 길 (동남지방통계청 지역통계과장)
이 진 원 (한국청소년활동진흥원 PBL지원부장)
최 재 원 (순천향대학교 경영학과 교수)
한 윤 선 (서울대학교 사회복지학과 교수)

청소년 빅데이터 체계 구축 및 활용방안 연구

인 쇄 2020년 12월 23일

발 행 2020년 12월 30일

발행처 한국청소년정책연구원
세종특별자치시 시청대로 370

발행인 김 현 철

등 록 1993. 10. 23 제 21-500호

인쇄처 경성문화사

사전 승인없이 보고서 내용의 무단전재·복제를 금함.
구독문의 : (044) 415-2125(학술정보관)

ISBN 979-11-5654-274-2 94330
979-11-5654-273-5 (세트)



연구보고 20-R08

청소년 빅데이터 체계 구축 및 활용방안 연구

 **한국청소년정책연구원**
National Youth Policy Institute

30147 세종특별자치시 시청대로370 세종국책연구단지
사회정책동(D동) 한국청소년정책연구원 6/7층
Social Policy Building, Sejong National Research Complex,
370, Sicheong-daero, Sejong-si, 30147, Korea
Tel. 82-44-415-2114 Fax. 82-44-415-2369 www.nypi.re.kr



9 791156 542742
ISBN 979-11-5654-274-2 94330
ISBN 979-11-5654-273-5 (세트)