

연구보고서(수시과제) 13

2016



여성가족 빅데이터 활용을 위한 기초연구

문유경 · 김영택 · 박건표 · 차지영



여성가족 빅데이터 활용을 위한 기초연구

연구책임자 : 문 유 경 (본원 선임연구위원)

공동연구자 : 김 영 택 (본원 연구위원)

박 건 표 (본원 연구원)

차 지 영 (이화여자대학교대 학교 교수)



한국여성정책연구원
Korean Women's Development Institute

이 보고서를 다음과 같이 인용해 주십시오.

문유경·김영택·박건표·차지영(2016). 여성가족 빅데이터 활용을 위한 기초연구.

서울: 한국여성정책연구원

한국여성정책연구원(구 한국여성개발원)은 국무총리실 산하 정부출연 연구기관으로, 한국의 여성정책에 대한 종합 연구를 통해 양성평등 실현, 여성의 사회참여 확대 및 복지 증진, 가족 그리고 국가 발전에 기여하는 여성정책 Think-Tank이다.

발 간 사

한국 정부는 2012년 “빅데이터 마스터 플랜”을 발표하고 데이터의 창조적 활용을 통한 스마트 강국으로서의 비전을 제시한 바 있습니다. 사회 현안에 대한 예방적 해결, 맞춤형 대국민 서비스를 가능하게 하는 새로운 분석 도구로서 빅데이터의 가능성을 강조하고 있습니다.

여성가족정책 분야는 정치, 경제, 사회 분야에 걸쳐 다양한 이슈를 갖고 있고, 대국민 서비스의 성격이 강한 만큼 빅데이터의 활용이 큰 도움이 될 것으로 예상됩니다.

본 연구는 본원에서 실시하는 최초의 빅데이터 연구로서 여성가족정책 연구에서 빅데이터를 향후 어떻게 활용할 것인지에 대한 기본방향 설정에 중점을 두고 연구를 수행하였습니다. 여성가족 빅데이터란 무엇을 의미하는지, 현재의 빅데이터 정책이 성인지적 관점에서 어떻게 개선되어야 하는지, 국제적 차원에서 성평등을 위한 빅데이터 활동은 무엇이 있는지 등을 중심으로 살펴보았습니다.

본 연구가 빅데이터를 활용한 여성가족 연구의 기본 방향을 제시하는 초석이 되기를 바라며, 향후 여성가족 연구의 새로운 지평이 열리기를 기대합니다.

끝으로 이 연구의 자문과 평가에 도움을 주신 원내외 여러 선생님들과 과제를 진행하여 주신 연구자들의 노고에 감사드립니다.

2016년 11월

한국여성정책연구원

원 장 이 명 선



연구요약

I. 서론

- 현대 사회에서 데이터의 중요성은 날로 높아져가고 있고, 기업은 물론 정책 결정자들과 연구자들도 데이터에 기반한 결정이 필수적이 되고 있음. 빅데이터는 이러한 데이터의 활용과 분석에서 핵심적인 개념으로 자리잡고 있음.
 - 여성가족 정책 수립과 연구 역시 지금까지 성인지 통계의 생산과 활용에만 의존하였던 것에서 탈피해 다양한 빅데이터를 이용한 연구결과를 도출함으로써 기존에 하지 못했던 사각지대를 커버할 수 있을 것으로 기대됨.

II. 여성가족 빅데이터의 정의와 범주

- 빅데이터 처리과정의 특성상 데이터가 초기부터 통계분석에 적합한 형태로 있기 보다는 복잡하고 다양한 분석과정을 통해 빅데이터로서의 역할을 하게 됨.
- 여성가족 빅데이터라는 개념은 사후적으로 여성과 가족에 대한 연구에서 사용된 빅데이터를 지칭하는 것이 적합하나 이러한 귀납법적인 접근방식으로는 빅데이터를 활성화 하는데는 한계가 있음.
- 본 연구에서는 이러한 문제점을 고려하여 여성가족 빅데이터를 단정적으로 정의하기 보다는 어떤 성격의 자료들이 여성가족 빅데이터로 활용가능성이 높은지를 탐색하였음.
 - 첫째, 여성가족 빅데이터로 활용가능성이 높은 자료들은 여성정책의 수행과 연구를 위해 지금까지 발생된 자료들이라 할 수 있음. 여성가족부와 한국여성정책연구원, 지방자치단체 여성가족 관련 부서, 지역의 여성정책연구소 등에서 축적된 자료들이 있음. 국가승인통계, 행정업무에서 발생한 통계, 각종 행정자료, 시설 및 센터에서의 실적과 상담자료, 복지사업의 대상자 관련 자료, 홈페이지와 해당 페이지 로그인 자료 등을 주제에 맞추어 결합

- 할 경우 빅데이터를 구축하여 새로운 연구결과를 얻을 수 있음.
- 둘째, 빅데이터의 생산과 활용에 성인지적 관점을 포함하거나, 성별 분리를 시도함으로써 발생하는 데이터를 예상할 수 있음. 예를 들어 서울시의 ‘서울시 정책지도’나 ‘우리마을가게 상권분석서비스’는 기초적인 행정자료와 지리정보 등을 이용하여 빅데이터를 분석한 획기적인 정책서비스임. 이와 같이 빅데이터의 구축에 있어서 성별차이를 보여주거나, 여성이슈 등을 고려한 내용을 포함한다면 여성가족 빅데이터의 범주로 활용할 수 있을 것임.
 - 셋째, 기존의 빅데이터 자료 유형들에 대해 성별로 분리하거나 여성 주제를 반영할 경우 여성가족 빅데이터의 활용이 가능할 수 있음. 예를 들어 신용/직불카드 금융자료 등이 현재 주요 빅데이터 자료로 활용되고 있음. 성별 이슈를 반영하여 신용불량자가 되는 원인이나 유형, 패턴 등을 신용카드 거래자료와 개인별 소득자료, 가구소득 자료 등을 연결하여 본다면 남성과 여성의 신용불량자 유입 경로의 유사점과 차이점 등을 분석할 수 있을 것이고, 성별로 적합한 예방책을 만들 수 있음.

III. 정부의 빅데이터 정책의 여성가족 관련성

- 정부의 빅데이터 정책을 보면 범정부 정책으로 현 정부는 2011년 교육과학기술부, 행정안전부, 지식경제부, 방송통신위원회, 국가과학기술위원회로 구성된 범정부 빅데이터 추진협의체를 구성하고 데이터의 창조적 활용을 통한 스마트 강국 실현을 비전으로 하는 추진전략과 목표를 수립함.
 - 추진과제로 6개 분야 빅데이터의 16개 대상과제를 제시하고 국민수혜, 유용성 등을 고려해 3개 과제를 우선추진과제로 선정하였음.
- 2013년 관계부처 합동으로 수립한 빅데이터 산업 발전전략은 3개 분야의 8개 추진과제를 수립하였음.
 - 이 전략은 국가차원의 빅데이터 ‘활용’을 촉진하고 빅데이터 ‘산업’을 육성하기 위한 전략적 지원이 필요한 시점임을 강조하고 있음.
- 주요 부처의 빅데이터 활용계획으로는 통계청, 국민안전처 등이 있음.
 - 통계청은 2015년 10월 빅데이터를 활용한 통계작성을 위해 통계데이터

허브국에 빅데이터통계과를 신설함. 빅데이터를 활용한 통계작성을 위한 제도적 기반을 구축하고 데이터 간 융·복합 사례를 발굴해 빅데이터 활용 확산에 나서고 있음. 데이터를 연계하고 분석결과를 확산하는 역할을 하는 데이터 허브 기능을 수행할 계획임을 밝히고 있음.

- 국민안전처는 2016년 전통적인 재난·안전 업무방식에 빅데이터를 접목 시키기 위해 재난예방 및 대응, 정책수립 지원, 활용기반 구축 등을 목표로 빅데이터 활용 종합계획을 수립함. 이 계획에는 2020년까지 5년 동안 12개의 신규과제와 5개의 기존과제를 포함 총 17개 과제가 있음.

- 2014년 미래창조과학부 주관 하에 부처의 과제 공모를 통해 최종 4개 부처에 대해 컨설팅을 진행함.

- 4개 부처의 하나인 여성가족부는 ‘13-’14년도 새일센터의 구인기업, 구직자, 취업자 특성과 일자리 미스매치 원인에 대한 분석을 시도함.

- 공공부문 빅데이터의 당면과제는 공공데이터의 개방과 개인정보보호의 충돌임.

- 개인정보 보호의 수집, 처리, 보유 절차가 엄격해짐에 따라 최소한 필요한 개인정보를 제외하고 폐기하는 과정에서 개인의 성별 식별 정보가 함께 삭제되는 경우 성별 분석이 어려워지는 문제가 있음.
- 개인정보 비식별화 과정의 필요성이 대두되고 있으며, 이 과정을 통해 개인의 식별은 불가능하지만 성, 연령 등의 속성은 유지되는 방향으로 빅데이터의 구축을 시도할 필요가 있음.

- 여성가족정책 빅데이터 연구를 위해서는 범정부 차원과 각 부처별 빅데이터 관련 세부과제를 통해 진행되는 사업의 경우 사업추진을 위해 활용하는 데이터의 성별구분 정보와 이를 통해 생산되는 데이터의 성별구분 식별 정보에 대한 검토가 필요함.

IV. 해외의 여성가족 빅데이터 활용 전략 및 사례

- 국제기구와 선진 국가들은 빅데이터를 이용한 정책개발과제 전략을 제시하여 강력하게 추진하고 있음. 미국의 ‘빅데이터 전략 수립 및 국가차원의 거버넌스

구축'이나 영국의 '정보공유·활용을 위한 데이터 개방중심의 정책' 등이 대표적인 사례임.

- Data2X는 빅데이터 분석을 기반으로 전세계 여성과 소녀들에게 보다 나은 삶을 제공하기 위하여 2012년에 형성된 협력적 플랫폼으로 UN의 우선계획(Initiative)으로 수행되고 있으며 William & Flora Hewlett 재단, Bill & Melinda Gates 재단 및 Clinton 재단이 지원하고 있음.
 - Data2X이 빅데이터 분석의 기준은 UN이 제시한 지속가능한 발전 목표(Sustainable Development Goals, SDGs)에 대한 성별영향평가로, 이를 위하여 UN 산하기관, 정부, 시민단체, 학계 등과 협력하여 젠더 데이터를 구축, 확장, 분석하여 효과적인 정책과 전략을 제공하고 의사결정을 제시함.
 - Data2X는 빅데이터 분석을 위하여 필요한 자료를 확인하면서 성별 데이터 차이(gender data gap) 변수를 확인하여 이를 건강, 교육, 경제기회, 정치참여, 안전의 다섯 가지 영역에서 28가지 변수로 분류하고 있음.
 - Data2X는 빅데이터 분석을 통해 전세계 여성건강을 도모하기 위한 소규모 프로젝트를 진행하고 있음(Data2X, 2015). 빅데이터를 이용한 성별분석 소규모 프로젝트는 휴대폰 이용 기록, 위성 데이터, 소셜미디어를 이용한 프로젝트 등이 있음.
 - Data2X가 갖는 중요성으로는 첫째, 여성의 성평등을 위해 데이터의 격차를 줄여야하고 이를 위해 빅데이터의 활용이 필요하다는 인식을 갖고 있는 프로그램이라는 점과 둘째, 직접 빅데이터를 생산하기 보다는 다른 기관과 협력체계를 구축하여 여성 빅데이터의 생산을 촉진하고 있다는 점임. 다른 기관과의 연계를 통하여 빅데이터를 효율적으로 생산하고 있는 점이 시사점임.
- UN의 빅데이터 글로벌 워킹 그룹은 상당히 많은 양의 데이터 사업을 소개하고 있음.
 - 이 사업 중 여성이나 가족과 관련된 내용으로는 '성별 분리된 재산권실태', '소셜 미디어를 활용한 직장에서 여성 차별 관련 패턴의 인지', '위생청결 관련 대중 인지에 대한 소셜 미디어 분석', 소셜 데이터를 활용한 '피임과 청소년 임신에 대한 태도 분석' 등이 있음.

V. 결론 및 정책제언

- 여성가족 빅데이터의 활용계획 수립
 - 여성가족 빅데이터의 활용에 대한 국가 전략과 데이터 거버넌스를 포함한 활용계획의 수립이 필요함.
 - 이 계획을 위해서는 사전에 여성가족 데이터의 빅데이터 활용 가능성과 빅데이터 활용 사업의 개발, 기술 인프라 구축 등에 대한 사전 연구가 필요함.
 - 이 계획에는 법, 제도와 데이터의 연계 공유에 대한 가이드라인, 활용을 위한 제도 개선 등이 포함되어야 할 것임.
- 여성가족정책 빅데이터의 원천자료 선별과 표준화
 - 여성가족정책의 실행과 연구에서 빅데이터의 이용을 활성화하기 위해서는 빅데이터의 가능성이 높은 자료를 선별하여 이들 자료의 축적이 선행되어야 함.
 - 여성가족부와 산하 기관 및 연구기관을 중심으로 여성가족정책 관련 조사 통계와 행정통계 및 사업실행 과정에서 시스템적으로 발생하는 정보를 총체적으로 점검하여 활용가능성과 필요성이 높은 자료들을 빅데이터로 활용할 수 있도록 하는 사전 작업이 필요함.
 - 선별된 자료들은 데이터 축적과 연계를 고려한 표준화 작업을 거쳐 빅데이터로의 활용성을 높임. 개인정보 보호를 준수하면서 자료 간의 연계가 가능하도록 하는 비식별화 과정에 대한 철저한 준비가 필요함.
- 여성가족정책 빅데이터의 DB 구축 및 운영
 - 여성가족에 대한 빅데이터 연구가 진행되고 데이터가 생산되기 시작하면 관련 정보를 제공하는 시스템의 필요성이 생겨남. 빅데이터의 성격상 모든 자료를 다 제공하기는 불가능하지만 빅데이터의 최종 분석결과 이전단계에서 활용할 수 있는 자료에 대한 수요가 커질 것.
 - 해당 빅데이터에서 활용하고자 하는 키워드를 선정하면 분석 가능한 형태의 데이터로 제공하는 시스템의 요구가 커질 것이며 이에 대한 준비가 필요함.
 - 한국언론진흥재단에서 서비스하고 있는 뉴스 빅데이터 분석 시스템인 BigKinds(빅카인즈)는 주요 언론사의 뉴스를 수집하는 시스템을 통해

형태소 분석, 개체명 분석, 네트워크 분석 등 기사를 분석하는 뉴스분석 시스템을 구축하고 있음.

- 빅 카인즈 프로(Big Kinds-Pro) 서비스는 더 전문화한 것으로 특정 뉴스에서 정보를 추출할 수 있도록 개인화 된 텍사노미 관리, 질의어 관리, 정보 추출 패턴 규칙 적용 기능을 제공하고 트렌드, 네트워크 분석 등 다양한 분석 도구를 통해 그 결과를 활용할 수 있음.

□ 정부 빅데이터 사업에 대한 성인지적 개선과 활용

- 국가승인통계의 신규 승인 시에는 통계법에 의해 특별한 사유가 없는 한 모든 개인차원의 통계는 성별 분리하여 조사, 발표하도록 되어 있음. 그러나 빅데이터는 공공부문에서 생산하더라도 성별 분리에 대한 의무조항이 없어, 성맹적(sex-blinded)으로 생산되고 있거나, 생산될 가능성이 높음.
- 빅데이터의 구축 시 개인정보가 있는 통계와 다른 정보의 연계 시에는 개인 비식별화 작업을 거치게 되는데 마찬가지로 성별 정보가 탈락할 가능성이 있으므로 이에 대한 대책이 필요함.
- 사용한 데이터가 성별구분이 불가능한 경우 왜 불가능한지에 대한 정보를 제공한다면 향후 관련 데이터 수집에서 성별구분 정보를 수집할 수 있는 방법과 데이터 수집 개선 방안의 기초자료로 제공될 수 있다는 점에서 큰 의의가 있음.

□ 여성가족 빅데이터 연구에 대한 정책지원

- 여성가족 빅데이터를 활성화하기 위한 기초연구에 대해 정책적 지원이 이루어져야 함.
- 연구주제로 첫째, 여성가족정책 관련 빅데이터 생산 및 연계방안 연구가 필요함. 빅데이터 연구가 필요한 여성가족정책 연구 주제를 발굴하고 빅데이터 생산방안을 검토하는 연구의 필요성이 제기됨. 생산의 개념은 새로운 데이터의 발굴 뿐 아니라 기존 데이터 연계를 포함하는 포괄적인 의미로 개별 부처 내에서 생산하는 데이터의 연계에서부터 부처 간 데이터의 연계, 공공데이터와 민간데이터의 연계까지 범위를 확장해 생산방안에 대한 종합적인 검토가 진행되어야 할 것임.
- 둘째, 여성가족정책 관련 국가승인통계 중 빅데이터와 연계하거나 대체가 가능한 조사통계에 대한 검토 및 발전방안을 수립하는 연구임. 정부 정책의

의사결정에서 데이터의 중요성은 점차 증대하고 있고 데이터 기반 환경에서 통계의 중요성은 더욱 강조 되고 있으나 조사통계의 경우 조사거부와 불응 등 응답률이 저하되는 문제로 인해 조사 환경이 악화되고 있음.

- 조사가 어려워지는 사회적 분위기 속에서 여성가족정책과 관련된 조사 통계의 어려움은 더욱 높아질 것이 자명해 보임. 특히, 경력단절, 다문화, 폭력, 성매매 등 민감성이 높은 조사가 상대적으로 많은 여성가족부의 상황을 고려할 때 데이터 연계를 통해 조사를 대체하거나 응답부담을 줄여 응답률을 높여나갈 수 있는 방안에 대한 검토가 요구된다고 할 수 있음.



목 차

I. 서 론	1
1. 연구의 목적 및 필요성	3
2. 연구내용 및 방법	4
II. 여성가족 빅데이터의 정의와 범주	5
1. 여성가족 빅데이터의 정의	7
2. 여성가족 빅데이터의 범주	11
III. 빅데이터 정책의 여성가족 관련성 분석	19
1. 정부의 빅데이터 정책 동향	21
가. 범정부 빅데이터 정책	21
나. 주요부처의 빅데이터 활용계획	24
다. 정부 주도의 공공 빅데이터 활용 사례	27
라. 공공부문 빅데이터 활용의 특징	34
2. 여성가족 빅데이터 연구 분야	36
가. 여성가족 빅데이터 연구 필요 분야	36
IV. 해외의 여성가족 빅데이터 활용전략 및 사례	41
1. 해외의 여성가족 빅데이터 활용 전략	43
2. 여성가족 빅데이터 활용 사례	56
가. 경제부문 활용사례	56
나. 안전부문 활용사례	57
다. 보건 부문 활용 사례	59

V. 결론 및 정책제언	61
1. 결론	63
가. 여성가족 빅데이터의 정의와 범주	63
나. 정부의 빅데이터 정책의 여성가족 관련성	64
다. 해외의 여성가족 빅데이터 활용 전략 및 사례	66
2. 정책제언	67
가. 활용계획의 수립	67
나. 여성가족정책 빅데이터의 원천자료 선별과 표준화	68
다. 여성가족정책 빅데이터의 DB 구축 및 운영	68
라. 정부 빅데이터 사업에 대한 성인지적 개선과 활용	69
마. 여성가족 빅데이터 연구에 대한 정책지원	69
 ■ 참고문헌	71
 ■ 부 록	73
부록 1. 경제 영역 부문 빅데이터 사업: UN 빅데이터 글로벌 워킹 그룹 (Big data global working group)	75
부록 2. 비 경제 영역 부문 빅데이터 사업: UN 빅데이터 글로벌 워킹 그룹 (Big data global working group)	78
 ■ Abstract	85

표 목 차

<표 II-1> 빅데이터 유형	9
<표 II-2> 여성가족부 국가승인통계 현황(2016)	11
<표 II-3> 빅데이터 원자료의 성별 분석 예상주제	18
<표 III-1> 빅데이터 산업 발전전략(2013) 세부추진과제	23
<표 III-2> 국민안전처 빅데이터 활용 신규과제 주요 내용(2016)	26
<표 III-3> 기관별 분석자료 형태 및 주요 분석내용	27
<표 III-4> e-새일센터 분석 수행 내용	29
<표 III-5> 국민 건강·질병표준DB 지표 종류	30
<표 IV-1> Data2X에서 제시한 성별 데이터 차이	50
<표 IV-2> Data2X 수행 빅데이터와 젠더 프로젝트	51
<표 IV-3> Data2X 협력기관, 역할, 담당영역 및 담당국가	53

그림 목 차

[그림 II-1] lot 자료의 분석과정	8
[그림 II-2] 여성가족부 정부 3.0 페이지	13
[그림 II-3] 여성가족부 공공데이터 개방 현황	14
[그림 II-4] 가정폭력 예방을 위한 빅데이터 활용 예상 시나리오	15
[그림 II-5] 서울시 빅데이터 활용 정책지도 홈페이지	16
[그림 II-6] 우리마을가게 상권분석서비스 홈페이지	17
[그림 III-1] 빅데이터 마스터플랜(2012)	21
[그림 IV-1] 빅데이터와 정책의 관계	44
[그림 IV-2] 미국 연방정부의 빅데이터 전략계획	46

I

서론

- | | |
|-----------------|---|
| 1. 연구의 목적 및 필요성 | 3 |
| 2. 연구내용 및 방법 | 4 |



1. 연구의 목적 및 필요성

이 연구의 목적은 여성가족 연구와 정책수립에 어떻게 빅데이터를 이용할 것인가에 대한 거시적 비전을 제공하는 일이다. 지금까지의 연구방법론은 양적 방법론과 질적 방법론으로 크게 나누어질 수 있으며, 양적 방법론은 통계분석을 통하여 이루어져 왔다. 통계분석의 기본적인 패러다임은 연구자가 제시하는 가설을 통계적인 검증을 통하여 기각 또는 수용함에 따라 기존의 이론을 유지, 확장, 수정하는 것이다. 빅데이터는 이러한 기존의 통계분석 패러다임을 갖고 분석하기에는 다른 성격을 갖고 있다. 통계자료의 수집, 분석방법, 결론의 타당화의 과정에서 빅데이터는 기존의 통계분석방법과 다른 측면을 갖고 있으며, 이러한 점이 기존의 통계자료 분석에서 얻을 수 없었던 새로운 사실을 발견하게 해 주고 있다. 본 연구는 빅데이터가 어떤 점에서 차별화되는 지를 찾아보고, 그 차별성을 이용하여 새롭게 할 수 있는 영역을 밝혀보는 일과 이를 수행하기 위해 필요한 데이터의 수집을 위해 필요한 일이 무엇인지 질문하고 있다. 구체적으로 첫째, 여성가족 연구와 정책에서 활용가능한 빅데이터에 대한 개념을 정의하려 한다. 일반적으로 교통안전 빅데이터, 보건의료 빅데이터 등 특정분야의 빅데이터에 대해서는 비교적 정의가 쉽고 분명하게 합의할 수 있지만 여성가족 분야의 빅데이터가 무엇인지에 대해서는 다양한 해석이 가능하며, 이에 대한 정리가 필요하다. 둘째, 지금까지 발표된 정부의 빅데이터 정책이 성인지적 관점에서 어떤 의미가 있는 지를 분석하려 한다. 정부의 빅데이터 정책의 주관부서는 정보, 통신, 기술 담당부서로 여성가족과 관련된 사회문제를 분석의 대상으로 선정하고는 있으나, 보다 근본적인 관점에서 빅데이터가 여성과 가족에게 미칠 수 있는 영향과 이에 대한 정책을 모색해보고자 한다.

셋째, 주요 선진국에서 여성정책에 빅데이터를 활용한 사례를 발굴하여 소개함으로써 향후 우리나라의 여성가족 빅데이터의 활용의 정책적 시사점을 제공하려 한다.

2. 연구내용 및 방법

주요 연구내용은 크게 세 부분으로 나누어진다. 첫째, 빅데이터와 여성가족 정책이 결합되는 새로운 개념은 어떤 내용을 포함하게 될 것인가에 대한 주제, 둘째, 공공부문 빅데이터 활성화 정책과 실제 활용 사례를 점검하여 여성가족의 관점에서 어떤 개선점이 필요한가에 대한 문제, 셋째, 외국의 여성가족과 관련된 빅데이터의 주요 활용전략과 사례에 대한 소개이다.

연구방법은 기존의 정책과 사례를 수집한 문헌연구, 빅데이터의 홈페이지 등에 대한 온라인 시스템 분석, 여성가족관련 통계와 행정자료에 대한 분석 등을 실시하였으며, 빅데이터 전문가를 초빙하여 연구진 워크숍을 개최하였고, 전문가 자문회의를 실시하여 연구의 전반적인 사항에 대한 자문을 구하였다.

II

여성가족 빅데이터의 정의와 범주

- | | |
|------------------|----|
| 1. 여성가족 빅데이터의 정의 | 7 |
| 2. 여성가족 빅데이터의 범주 | 11 |



1. 여성가족 빅데이터의 정의

빅데이터가 무엇을 의미하는 가는 연구자의 입장에 따라 다양하다. 채승병·안신현·전상인은 “빅데이터는 단지 거대하기 보다 순환속도가 매우 빨라서 기존 방식으로서는 관리, 분석이 어려운 데이터를 의미”하는 것으로 정의하고 있다(채승병, 안신현, 전상인(2012):5). 그는 “빅데이터를 규정하는 요소로 3V”를 들고 있는데 거대한 크기(Volume), 다양한 형태(Variety), 빠른 속도(Velocity)이다(채승병, 안신현, 전상인(2012):5). 첫째, 빅데이터의 특성으로 데이터 크기를 들고 있다. 일반 기업에서도 테라바이트(TB)에서 페타바이트(PB)¹⁾ 규모의 데이터를 다루는 것이 가능해지면서 빅데이터가 가능해 진다는 점이다. 둘째, 빅데이터의 특징으로 데이터의 형식의 이질성을 들고 있다. 동일한 형식이 아닌 다양한 형식 예를 들어 텍스트 메시지, 영상, 음성 등의 자료를 결합하여 빅데이터를 생성함으로써 기존의 자료분석과 다른 차원의 분석이 가능하다는 것이다. 셋째, 빠른 속도이다. 데이터의 생성 후 유통되고 활용되기까지의 소요시간이 단축되고 있다. 전형적인 자료분석이 기획, 조사, 수집, 분석을 일련의 과정으로 진행해온 것에 비해 동시에 할 수 있게 된 점을 들고 있다.

정용찬은 빅데이터 환경의 특징으로 데이터의 양이 폭증하였고, 데이터의 종류도 다양해졌다는 점과 함께 관련 장비와 소프트웨어 측면에서 분산병렬 처리 기술, 클라우드 컴퓨팅을 활용해 효율적인 시스템 구축이 가능해 졌다고 보고 있다(정용찬(2012):6).

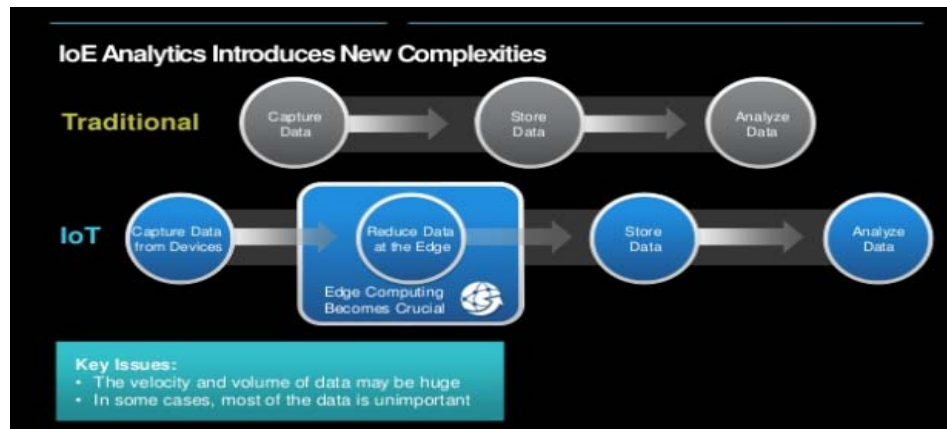
이러한 세 가지 특성과 빅데이터 환경의 특징은 빅데이터를 잘 설명하고 있지만 여전히 의문점이 남는다. 어떤 것이 빅데이터이고, 어떤 것이 아닌가에 대한 기준이 모호하다는 점이다. 크고 다양하고 빠른 데이터를 모두 빅데이터로 간주하기에는 현실적으로 너무나 많은 자료가 도처에 널려 있다.

라지브 닐스(Rajiv Niles)는 기존의 자료와 달리 빅데이터의 하나인 사물 인터넷의 분석과정을 설명하면서 자료를 특정 목적으로 축약하는 과정이 매우 중요함을 밝히고 있다. 기존의 방법은 목적을 갖고 자료를 수집한 후

1) 1페타바이트=1,000테라바이트=100만GB=10억MB, 1엑사바이트(EB)=1조MB, 1제타바이트(ZB)=1000조MB

8 ●●● 여성가족 빅데이터 활용을 위한 기초연구

그것을 분석했던 반면, IoT의 분석방법은 주어진 자료 안에서 취사선택의 과정을 거친 후에 그 자료를 해석하는 것이 더욱 중요해졌다고 주장한다.



출처: Rajiv Niles(2015), “Data & Analysis in the Internet of Everything” 매일경제신문, 미래창조과학부. 「2015년 빅데이터컨벤션 및 성과보고회」 발표자료.

[그림 II-1] IoT 자료의 분석과정

라지브 닐스의 주장처럼 빅데이터는 자료의 취급과정이 더욱 중요성을 갖게 되며, 연구자는 이러한 관점을 반영하여 빅데이터의 정의의 하나로 사후적 관점을 도입하려 한다. 빅데이터의 성격상 모든 것이 자료로 활용될 수 있는 것이 아니라 많은 자료의 더미 속에서 필요한 내용을 추출하고 그것을 해석할 수 있는 과정이 중요하다. 즉, 빅데이터는 처음부터 정의할 수 있는 것이 아니라, 어떤 주제에 대한 분석을 위해, 다른 다양한 자료를 결합하여 새로운 형식의 자료를 생성하여 결과를 얻었을 경우 빅데이터를 활용하였다고 결과적으로 말하는 것이 현재로서는 정확한 표현으로 생각된다.

이러한 관점은 통계학과 데이터 마이닝의 비교를 통해서도 잘 나타난다.

통계학과 데이터 마이닝의 차이를 살펴보면 통계학은 비교적 크지 않은 실험데이터를 대상으로 하는데 반해 데이터 마이닝은 비계획적으로 축적된 대용량의 데이터를 대상으로 한다. 통계학이 추정(estimation)과 검정(testing)이라는 이론을 중시하는 특징을 가졌다면 데이터 마이닝은 이해하기 쉬운 예측모형의 도출에 주목한다. 즉, 데이터 마이닝은 기업 활동 과정에서

Ⅱ. 여성가족 빅데이터의 정의와 범주 ●●● 9

자연스럽게 축적된 대량의 데이터를 분석해 기업 경영에 필요한 가치 있는 정보를 추출하기 위해 사용된다. 이러한 이유로 데이터 마이닝을 “규모, 속도, 그리고 단순성의 통계학(statistics at scale, speed, and simplicity)”이라 부른다(정용찬(2013):34).

지금까지 어떤 자료가 빅데이터로 분석되었는지를 정리하면 주요 유형으로 거래자료, 추적장치자료, 센서자료, 온라인 활동자료, 의견자료, 행정자료 등이 있다(이지영(2015), “빅데이터의 국가통계 활용을 위한 기초연구” 53쪽). <표 II-1>에서 나타나듯이 빅데이터는 어떤 원천자료를 어떻게 결합하는가에 따라 무궁무진한 형태로 생성될 수가 있다. 따라서 빅데이터의 정의에서 원천자료 자체를 빅데이터로 간주하기는 어려우며, 분석을 위해 다양한 원천자료를 선정, 가공, 결합 등의 과정을 거친 자료를 빅데이터로 볼 수 있을 것이다.

〈표 II-1〉 빅데이터 유형

원 천 자 료	빅데이터 원자료	활 용 사 례
거래자료 (Transaction data)	신용/직불카드 및 금융 거래 자료, 유통업체 자료	(뉴질랜드 통계청) 거래통계 (한국석유공사) 물가통계
추적장치자료 (Tracking device data)	통신사의 모바일 위치자료 앱기반/상업용차량 GPS자료	(네덜란드/유럽/뉴질랜드 통계청)유동인구통계 (부산광역시)서비스인구통계
센서자료 (Censor data)	위성사진, 도로/기상/ 전력센서	(네덜란드 통계청) 교통통계 (호주 통계청) 통업통계 (아일랜드 통계청) 전력통계 (기상청) 기상정보융합서비스
온라인 활동자료 (Online behavior data)	온라인 가격정보 온라인 검색, 등록 정보	(네덜란드/유럽 통계청) 물가통계 (이탈리아 통계청) 기업 ICT관련 통계
의견자료 (Opinion data)	트위터 등 공개 SNS 자료	(네덜란드 통계청) 심리상태 (UN Global Pulse) 고용통계
행정자료 (Administrative data)	공공·행정자료(행정, 납세, 복지, 보건, 주택 등)	(호주/덴마크 통계청) 인구통계 (한국교육개발원)취업통계

출처: 이지영(2015). “빅데이터의 국가통계 활용을 위한 기초연구”. 통계개발원. 『2015년 상반기 연구보고서 제Ⅱ권』. 53쪽.

10 ●●● 여성가족 빅데이터 활용을 위한 기초연구

이와 같이 정의될 수 있는 빅데이터를 여성가족과 연결하여 ‘여성가족 빅데이터’라는 용어를 만들어본다면 이에 대한 정의는 몇 가지가 가능하다. 우선 여성과 가족 개념의 분리이다. 두 개념의 상관성이 높기 하지만 서로 결합되어 새로운 의미를 생성하는 것은 아니므로 여성가족 빅데이터는 ‘여성과 가족 빅데이터’로 이해해야 할 것이다.

먼저 ‘여성과 가족을 분석대상’으로 한 빅데이터로 해석할 수 있다. 이러한 해석에서 발생하는 문제는 분석의 목적과 분석주체이다. 이미 많은 기업들이 마케팅의 일환으로 소비자로서의 여성에 대한 빅데이터를 활용하고 있다. 인터넷의 SNS 등 온라인 자료, 카드 자료, 모바일 자료 등을 이용하여 성, 연령, 계층, 지역 등 전형적인 분석과 장소와 시점의 흐름이나 패턴의 유형화 등을 결합한 빅데이터를 생산하고 있다. 이와 같이 특정 목적을 위하여 여성가족을 분석대상으로 생산된 빅데이터를 우선 빅데이터로 간주할 수 있을 것이다. 그러나 이러한 분석대상으로만 제한하기 보다는 여성과 가족을 위한 빅데이터를 생각할 수 있다. 즉 성평등과 여성의 지위향상, 성별 차이 등의 연구와 정책수립을 위한 빅데이터를 고려할 수 있으며, 이 범주는 전자보다 훨씬 광범위하게 포함할 수 있을 것이다. 가족 역시 가족의 생성과 소멸, 가족구성원 간의 관계 등에 대한 연구와 정책을 위한 빅데이터를 고려할 수 있으며, 이와 같이 여성과 가족의 연구와 정책수립에 필요한 빅데이터를 ‘여성가족 빅데이터’의 정의에 포함해야 빅데이터를 활용한 연구가 사회적 의의를 갖게 될 것이다.

2. 여성가족 빅데이터의 범주

여성가족 빅데이터로 활용가능한 첫 번째 범주는 여성가족 정책과 연구의 수행과정에서 발생하는 자료들을 상정할 수 있다. 여성가족부와 지방자치단체의 관련 부서, 한국여성정책연구원을 비롯한 시도별 여성연구원 등에서 생산되는 자료들의 다양한 결합은 여성가족 빅데이터의 원천자료로 활용 가능할 것이다.

여성가족부에서 생산가능 한 빅데이터의 원천자료들을 보면 15종의 국가승인통계와 업무수행에서 발생하는 행정통계, 산하 기관과 센터 운영과 관련된 행정자료, 홈페이지와 관련된 온라인 자료 등이 가능하다. <표 II-2>는 여성가족부 국가승인통계 현황이다. 가정폭력실태조사 등 정책수행에 필요한 자료의 수집을 위해 조사가 13종이며, 현황에 관련된 통계가 2종이다. 국가승인통계는 그 자체만으로도 정책지원의 기능을 하고 있지만 동일 주제의 행정자료나 기타 온라인 자료들과 연계하여 새로운 빅데이터를 생산할 경우 구심점 역할을 할 수 있다.

〈표 II-2〉 여성가족부 국가승인통계 현황(2016)

번호	통계명	작성 형태	통계 분야
1	가정폭력 실태조사 ²⁾	조사	복지
2	가족 실태조사 ³⁾	조사	인구
3	경력단절 여성등의 경제활동 실태조사 ⁴⁾	조사	복지
4	국민 다문화 수용성 조사 ⁵⁾	조사	복지
5	국제결혼 중개 실태조사 ⁶⁾	조사	사회
6	기업 및 공공기관의 가족친화 수준조사 ⁷⁾	조사	복지
7	성폭력 실태조사 ⁸⁾	조사	복지
8	성희롱 실태조사 ⁹⁾	조사	사회
9	양성평등실태조사 ¹⁰⁾	조사	사회
10	전국 다문화가족 실태조사 ¹¹⁾	조사	사회
11	중앙행정기관 및 지방자치단체 소속 위원회 여성참여현황 ¹²⁾	보고	사회
12	청소년 상담지원 현황 ¹³⁾	보고	사회

12 ●●● 여성가족 빅데이터 활용을 위한 기초연구

번호	통계명	작성 형태	통계 분야
13	청소년 매체이용 및 유해환경 실태조사 ¹⁴⁾	조사	사회
14	청소년 종합 실태조사 ¹⁵⁾	조사	사회
15	한부모 가족 실태조사 ¹⁶⁾	조사	복지

출처: 여성가족부 홈페이지 정책안내>정책자료>통계자료>‘여성가족부 국가승인통계 사전
공표’의 표를 재구성하였음.
http://www.mogef.go.kr/korea/view/policy/policy02_05a.jsp?func=view¤tPage=0&key_type=&key=&search_start_date=&search_end_date=&class_id=0&idx=704218.2016.11.22.)

- 2) 가정폭력에 관한 일반인의 인식과 피해 실태, 발생요인, 발생유형, 피해자 욕구, 지원시설의 서비스와 운영 상황 등 가정폭력 전반에 대한 종합적 실태조사
- 3) 가족구성원의 일반 특성(성별·연령·학력·혼인상태·취업상태·건강상태 등), 경제상태(소득·지출·자산 등), 가치관, 가족행태, 가족관계, 생활양식, 가족 문제, 서비스 욕구 등에 대한 실태조사
- 4) 임신·출산·육아 등의 이유로 경력이 단절된 여성 등의 경제활동 실태를 파악하여 재취업지원 등 정책수립에 필요한 기초자료 확보
- 5) 우리 국민의 다문화 사회에 대한 인식수준을 파악하고, 정책 관련 변수들을 진단하여 다문화 정책수립의 기초 자료로 활용하기 위한 조사
- 6) 국제결혼중개업체 현황, 이용자 현황(연령별·학력별·직업별·소득별 이용자 수 및 출신국별·연령별·학력별 상대방수, 결혼생활 만족도 등) 피해현황(중개업체 이용자의 피해유형 및 유형별 피해 세부현황 등) 등을 조사
- 7) 전국 227개 시·군·구의 행정 인프라, 돌봄 서비스 인프라 (아동, 청소년, 노인 등)와 여가·문화 시설 및 지역 주민대상 욕구·만족도 조사
- 8) 성폭력에 관한 일반인의 인식과 피해 실태, 발생요인, 발생유형, 피해자 욕구, 지원시설의 서비스와 운영 상황 등 성폭력 전반에 대한 종합적 실태조사
- 9) 직장 내 성평등 문화, 성희롱 관련업무 운영현황, 성희롱 예방교육, 성희롱 실태 및 현황, 업무담당자의 성희롱 심각성 인지 및 업무의 어려움 등 조사
- 10) 양성평등에 대한 국민의 인식 및 태도, 정책 수요 등에 대한 실태를 파악하여 양성평등정책 기본계획 수립* 등 정책 추진에 활용
- 11) 전국 다문화가족의 현황 및 실태를 파악하여 정책수립에 필요한 기초 통계자료 구축
- 12) 위원회 설치근거(조례 등), 여성참여율(당연직 및 위촉직), 위촉직 여성위원 명단 조사
- 13) 전국 청소년상담복지센터 및 한국청소년상담 복지개발원 등을 통한 상담 내용을 학업 및 진로, 대인관계, 일탈 및 비행, 정신건강, 컴퓨터 및 인터넷 사용, 성격, 가족관계 등을 초등학교, 중학교, 고등학교, 대학생, 기타로 분석 정리
- 14) 청소년 매체 이용 실태 조사와 청소년유해환경 접촉 종합실태조사 통합 조사
- 15) 가정생활 및 양육정보, 생활환경, 학교생활, 건강·신체, 인권·참여, 활동·여가, 진로·아르바이트 등 종합적인 실태 파악
- 16) 한부모가족의 소득 및 지출, 전 배우자와의 관계 및 양육비, 경제활동, 아이돌봄, 건강, 주거 환경, 생활세계 및 사회적 지지망, 지원정책 수요 등을 조사

II. 여성가족 빅데이터의 정의와 범주 ●●● 13

현 정부의 국정과제 중 하나인 ‘정부 3.0’은 정부에서 생산하고 있는 공공데이터의 개방을 통한 부처 간의 협력과 민간과의 소통을 목표로 하고 있다. 정부 각 부처는 ‘정부 3.0’과제에 의해 각종 행정 자료를 공개하고 있다. 여성가족부의 홈페이지에도 ‘정부3.0’이 메인 메뉴로 자리 잡고 있으며, 정부 3.0의 주요 내용은 에 집적되어 있다. 여기에는 정보공개, 법령정보, 정책설명제 등의 내용이 게재되어 있다.

정보공개제도안내

HOME > 정보공개 > 정보공개 > 정보공개제도안내

정보공개제도안내

정보공개제도란?

공공기관이 직무상 작성 또는 취득하여 관리하고 있는 정보를 수요자인 국민의 청구에 의하여 공개하거나, 중요 정보를 사전에 국민에게 제공함으로써 국민의 알권리를 보장하고, 국정운영에 대한 국민참여와 행정의 투명성을 제고할 수 있도록 하기 위한 제도입니다.

청구절차 및 대상정보

구분	내용	
정보공개 청구권자	모든 국민	모든 국민은 청구인 본인 또는 그 대리인을 통하여 공공기관에 정보공개를 청구할 권리를 가지고 있습니다.
	법인 단체	법인과 단체는 대표자의 이름으로 공공기관에 정보공개를 청구할 권리를 가지고 있습니다.
	외국인	<ul style="list-style-type: none"> 청구권이 인정되는 외국인의 범위는 다음과 같습니다. <ul style="list-style-type: none"> 국内に 영주권을 두고 거주하는 자 학술, 연구를 위하여 일시적으로 체류하는 자 국내에 사무소를 둔 법인 또는 단체
대상정보	<ul style="list-style-type: none"> 본부 및 소속기관이 직무상 작성 또는 취득하여 관리하고 있는 문서, 도면, 사진, 필름, 테이프, 슬라이드 및 컴퓨터에 의하여 처리되는 매체 등에 기록된 사항 (법 제2조 제1호) 알려진 공고, 고시, 예고 또는 등본, 초본 그 밖의 사본이 고부대상이 되는 행정정보에 대하여는 적용하지 않음. 각 기관의 자료실 등에서 국민에게 열람 또는 대출되는 도서, 간행물 등을 대상정보에서 제외함 국민 생활에 중대한 영향을 미치는 정책 및 행정에 관한 정보 <ul style="list-style-type: none"> 대규모 예산이 투입되는 공공사업에 관한 정보 	

출처: http://www.mogef.go.kr/korea/view/policy/policy04_01_01a.jsp.

[그림 II-2] 여성가족부 정부 3.0 페이지

정보공개의 세부 메뉴 중의 하나인 ‘공공데이터 공개’에는 ‘여성인재 DB등록현황’ 등 39개의 파일데이터와 5종의 오픈 API가 탑재되어 있어 빅데이터의 원천자료로 활용가능성이 있다.

14 ●● 여성가족 빅데이터 활용을 위한 기초연구

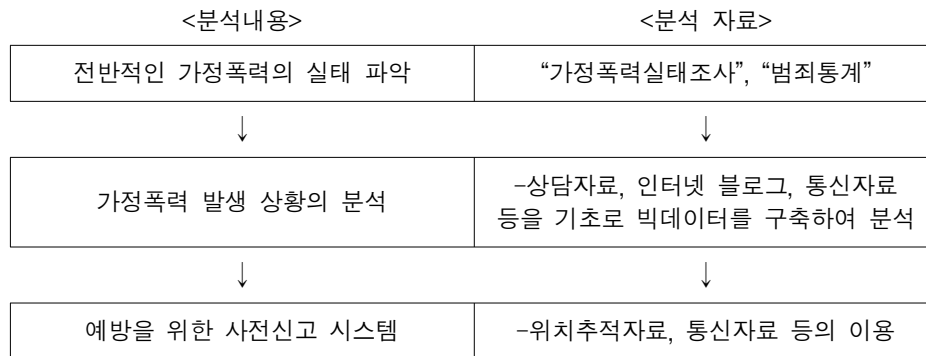
전체(44)	파일데이터(39)	오픈API(5)	표준데이터(0)
파일데이터 39건을 찾았습니다.			
<div> <div>파일데이터 [39건]</div> <div>정확도순 ▾ 날짜순 ▾ 제목순 ▾ 조회순 ▾ 다운로드순 ▾</div> </div>			
여성인재DB 등록 현황 조회수 : 323 다운로드수 : 13 수정일 : 2016.11.25 기관 : 여성가족부 서비스유형 : 다운로드 여성인재DB 연도별, 직종별 현황 CSV			사회복지 사회복지
여대생커리어개발센터 지정 현황 조회수 : 212 다운로드수 : 16 수정일 : 2016.11.21 기관 : 여성가족부 서비스유형 : 다운로드 여대생커리어개발센터 지정 현황(2010년 이후 27개 센터) CSV			사회복지 사회복지
지역별 성평등 수준 분석 연구 조회수 : 364 다운로드수 : 128 수정일 : 2016.10.24 기관 : 여성가족부 서비스유형 : 다운로드 시도별 성평등수준과 변화 추이 등을 측정, 분석한 성평등보고서 HWP			사회복지 사회복지
한국의 성평등 보고서 조회수 : 447 다운로드수 : 200 수정일 : 2016.10.21 기관 : 여성가족부 서비스유형 : 다운로드 국가성평등지수와 지표를 구축하여, 이를 기초로 우리나라의 성평등수준과 변화 추이와 그 원인 등을 분석 HWP			공공행정 공공행정

출처: http://www.mogef.go.kr/korea/view/policy/policy04_11.jsp. 2016.11.22.

[그림 II-3] 여성가족부 공공데이터 개방 현황

마지막으로 빅데이터로서 활용가능성이 높은 여성가족부 자료로는 온라인 자료가 있다. 여성가족부 홈페이지 로그인 자료, 게시판 등의 의견 자료, SNS에 나타나는 여성가족부와 여성가족정책 관련 내용에 대한 자료 등이 빅데이터로서의 활용 가능성이 있다. 홈페이지 로그인 자료는 특히 특정 주제의 통계자료 및 행정자료 등과 결합하여 빅데이터로 활용할 수 있다. 예를 들어 ‘성폭력’ 예방 정책을 수립할 때 국가통계인 “가정폭력실태조사”의 통계자료와 홈페이지의 “가정폭력 예방” 페이지의 로그인 자료(누가 무엇을 언제 조회하였는지), 행정자료인 “가정폭력 상담 시설” 이용자료, 인터넷 블로그 등을 결합할 경우 새로운 차원의 빅데이터 분석이 가능할 것이다.

Ⅱ. 여성가족 빅데이터의 정의와 범주 ●●● 15



[그림 Ⅱ-4] 가정폭력 예방을 위한 빅데이터 활용 예상 시나리오

여성가족부의 국가승인통계와 홈페이지에 공개된 공공데이터와 정보공개된 내용 외에도 산하 기관의 이용자 관련 자료, 행사 참가자 자료, 사업 대상자 자료 등도 향후 빅데이터로 활용 가능성이 크다고 볼 수 있다.

여성가족부 외에 한국여성정책연구원의 연구수행에서 발생하는 자료도 여성가족 빅데이터의 활용가능성이 크다, “여성가족패널조사” 등 국가승인 통계와 성인지 통계시스템에 탑재된 통계 등 공식통계와, 성별영향분석평가 및 성인지 예산 사업 등 대형 사업에서 발생하는 행정자료, 홈페이지 로그인 자료 등도 주제에 맞추어 결합할 경우 ‘여성가족 빅데이터’로서의 활용가능성이 크다. 그 외에 지방자치단체의 여성가족 관련 부처, 지방의 여성가족정책연구원 등에서 발생하고 있는 각종 통계와 행정자료 등도 ‘여성가족 빅데이터’의 원천자료로 활용 가능할 것이다.

두 번째 범주는 빅데이터의 생산과 활용에 성인지적 관점을 포함하거나, 성별 분리를 시도함으로써 발생하는 데이터를 예상할 수 있다. 기존의 성인지적 통계 정책에서는 개별 차원의 모든 통계를 성별로 분리하거나 성인지적 이슈가 있는 통계를 생산하는 것을 주요 전략으로 활용하여 왔다. 빅데이터 역시 여성과 가족과 관련된 이슈를 발굴하여 그 이슈를 해결하기 위해 생산된 빅데이터를 여성가족 빅데이터로 범주화할 수 있으며, 동시에 기존의 빅데이터 분석 결과를 검토하여 성별 차이가 예상되는 경우 재분석할 경우 여성가족 빅데이터로 범주화할 수 있다.

16 ●● 여성가족 빅데이터 활용을 위한 기초연구

서울시는 빅데이터를 활용하여 311종의 “정책지도(http://gis.seoul.go.kr)”를 공개하였다. 서울시에서 보유한 기초적 행정자료와 지리정보시스템 기법을 결합한 공간정보 빅데이터로 ▲안전한 서울(안전, 재난) ▲따뜻한 서울(복지, 주거) ▲꿈꾸는 서울(경제, 일자리) ▲숨쉬는 서울(환경, 교통)의 4개 목표를 갖고 제작되었다. 이 정책지도의 경우 빅데이터를 행정에 적극적으로 활용한 모범사례라고 볼 수 있다. 그러나 성인지적 관점에서 보면 성별 차이를 드러내거나, 여성의 관점에서 필요한 주제들 예를 들어 안전 등에 대한 지도는 찾기 힘들다. 이러한 공공정책에서 빅데이터의 생산 시 여성정책의 수요를 적극적으로 반영한다면 여성가족 빅데이터의 또 하나의 범주가 생겨날 수 있다.



출처: <http://gis.seoul.go.kr>

[그림 II -5] 서울시 빅데이터 활용 정책지도 홈페이지

Ⅱ. 여성가족 빅데이터의 정의와 범주 ●● 17

또 다른 예로 서울시에서 발표한 “우리마을가게 상권분석서비스 (<http://golmok.seoul.go.kr>)”를 들 수 있다. 이것은 서울시내 대로변의 뒷골목 등 1008개 영세한 골목상권 총 43개 생활밀착업종에 대한 빅데이터를 분석해 신규창업 위험도나 지역 상권분석 정보 등을 검색할 수 있는 서비스다. 이러한 서비스를 기존의 ‘사업체 실태조사’ 등을 이용해 사업체 대표자의 성별 업종별 분포와 연결하여 분석한다면 성별 사업체 운영실태의 차이점을 보다 세밀하게 분석할 수 있을 것이다.



출처: <http://golmok.seoul.go.kr>

[그림 Ⅱ-6] 우리마을가게 상권분석서비스 홈페이지

여성가족 빅데이터의 세 번째 범주로는 현재 빅데이터 원자료로 활용되고 있는 것들을 중심으로 성별의 차이를 모색해볼 수 있다. 예를 들어 신용불량자가 되는 원인이나 유형, 패턴 등을 신용카드 거래자료와 개인별 소득자료, 가구소득 자료 등을 연결하여 본다면 남성과 여성의 신용불량자 유입경로의 유사점과 차이점등을 분석할 수 있을 것이고, 성별로 적합한 예방책을 만들 수 있다.

18 ●● 여성가족 빅데이터 활용을 위한 기초연구

〈표 Ⅱ-3〉 빅데이터 원자료의 성별 분석 예상주제

원 천 자 료	빅데이터 원자료		여성 빅데이터로 활용 가능한 주제
거래자료 (Transaction data)	신용/직불카드 및 금융 거래 자료, 유통업체 자료	→	예)신용/직불카드의 성별 사용현황과 사용금액의 차이
추적장치자료 (Tracking device data)	통신사의 모바일 위치자료 앱기반/상업용차량 GPS자료	→	예) 성별 지리적 이동패턴
센서자료 (Sensor data)	위성사진, 도로/기상/ 전력센서	→	예) CCTV의 장소별 시간대별, 성별 출현자수
온라인 활동자료 (Online behavior data)	온라인 가격정보 온라인 검색, 등록 정보	→	예)온라인 검색, 등록활용의 성별 차이 온라인쇼핑의 행태차이
의견자료 (Opinion data)	트위터 등 공개 SNS 자료	→	예)주제별 SNS자료의 성별 사용언어의 차이
행정자료 (Administrative data)	공공·행정자료(행정, 납세, 복지, 보건, 주택 등)	→	예)성별 납세 현황과 복지 상황등의 연결

출처: 이지영(2015). “빅데이터의 국가통계 활용을 위한 기초연구”. 통계개발원. 「2015년 상반기 연구보고서 제Ⅱ권」. 53쪽을 기초로 재작성

III

빅데이터 정책의 여성가족 관련성 분석

- | | |
|--------------------|----|
| 1. 정부의 빅데이터 정책 동향 | 21 |
| 2. 여성가족 빅데이터 연구 분야 | 36 |

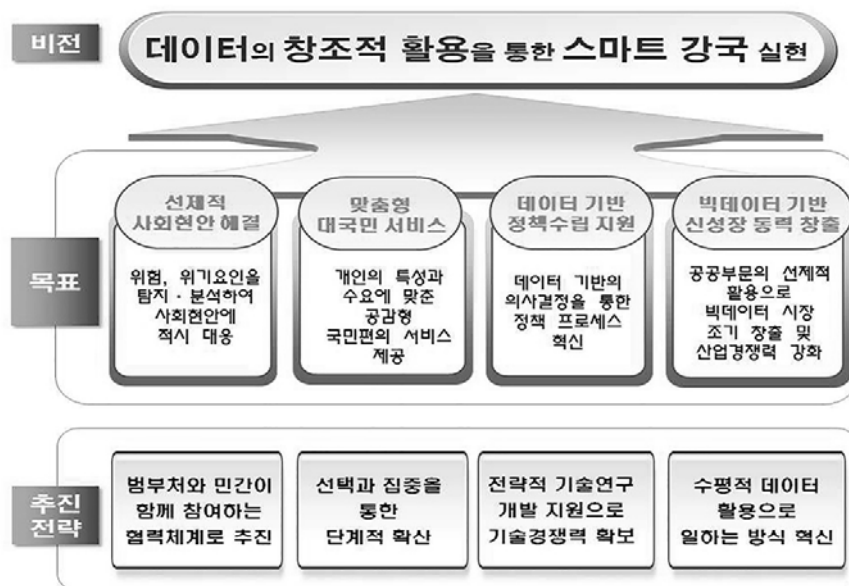


1. 정부의 빅데이터 정책 동향

가. 범정부 빅데이터 정책

1) 빅데이터 마스터플랜¹⁷⁾

2012년 11월 교육과학기술부, 행정안전부, 지식경제부, 방송통신위원회, 국가과학기술위원회로 구성된 범정부 빅데이터 추진협의체는 데이터의 창조적 활용을 통한 스마트 강국 실현을 비전으로 하는 추진전략과 목표를 수립했다. 추진전략으로는 범부처와 민간이 함께 참여하는 협력체계, 선택과 집중을 통한 단계적 확산, 전략적 기술연구개발 지원으로 기술경쟁력 확보, 수평적 데이터 활용으로 일하는 방식 혁신을 선정했다.



출처: 교육과학기술부, 행정안전부, 지식경제부, 방송통신위원회, 국가과학기술위원회, 스마트 국가 구현을 위한 빅데이터 마스터플랜. 17쪽

[그림 Ⅲ-1] 빅데이터 마스터플랜(2012)

17) 2012.11, 교육과학기술부, 행정안전부, 지식경제부, 방송통신위원회, 국가과학기술위원회, 스마트 국가 구현을 위한 빅데이터 마스터플랜

22 ●●● 여성가족 빅데이터 활용을 위한 기초연구

추진과제로는 국민적 관심이 큰 6개 분야 빅데이터 대상과제 16개를 제시해 국민 수혜, 유용성 등을 고려해 3개 과제를 우선추진하고 단계적으로 확산하는 방향으로 추진했다. 우선추진 활용과제로 선정된 과제를 살펴보면 먼저 범죄발생 장소·시간 예측을 통한 범죄발생 최소화가 있다. 경찰청에서는 수사, 범죄분석 등을 위해 범죄정보관리시스템과 지리적 프로파일링 시스템, 과학적 범죄분석시스템 등을 사용 중이나 타 기관과의 연계와 활용이 미흡해 지역별 범죄위험 시간 예측과 실시간 활용은 제한적이다. 사업추진 모델은 정형 데이터와 비정형 데이터를 연계해 장소별·시간대별 범죄 발생 가능성을 도출하는 범죄발생 예측서비스를 제공하는 것으로 정형데이터는 지역별·시간별 과거 범죄발생 내역, 순찰경로 등이고 비정형 데이터는 주민신고, CCTV의 범죄 행동분석 등이다.

다음으로 선정된 우선과제는 예측 기반의 자연재해 조기 감지 대응 사업이다. 자연재해의 경우 집중강우, 하천범람 등을 분석하고 선제적으로 대응하기 위해 재해상황 분석·판단시스템을 구축하였으나 수집된 정형 데이터의 부정확성, 빅데이터 분석 처리의 복잡성 등으로 인해 정확한 예측이 어렵다는 한계가 있었다. 그래서 집중호우에 따른 내수침수, 산사태 등을 실시간으로 조기에 감지해 재난 피해지역을 사전에 예측하는 서비스를 제공하는 것을 목표로 한다. 기상정보, 수문정보, 절개지, 지역별 인구정보 등 정형데이터를 활용하고, CCTV, 인공위성 자료, 소셜 데이터, 주민신고 등 비정형데이터를 활용한다.

마지막으로 주민참여형 교통사고 감소체계 구축 사업은 현재 교통사고 정보관리시스템을 통한 교통사고 이력, 자동차 보험, 차량등록, 공간정보, 의료 DB 등 정형데이터를 활용하고 있으나 지역별 교통안전 환경에 익숙한 주민참여 방식의 교통안전 개선 프로그램 운영이 실질적인 교통사고 감소로 이어진다는 일본 카마가야 시의 사례를 통해 필요성이 대두되었다. 교통안전 관련자료 및 주민참여 개선의견 등을 이용해 분석모델을 구축하고 이를 정책에 활용하는 교통사고 발생 최소화를 목표로 하고 있다. 사용되는 정형 데이터로는 자동차보험, 차량등록, GIS, 지능형 교통정보, 의료정보 등이 있고 비정형 데이터로는 주민 개선의견, 게시판, 교통방송 등의 교통사고 위치 등이 있다. 데이터를 결합해 신호등, 과속 방지턱 설치 등의 개선방안을

Ⅲ. 빅데이터 정책의 여성가족 관련성 분석 ••• 23

도출하고 정형데이터는 발생지점별 사고규모, 피해액 등을 비정형데이터는 사고유형, 시간대별 교통사고 발생지점 발견으로 개선을 통해 교통사고가 줄어들 것이라 기대하고 있다.

2) 빅데이터 산업 발전전략(관계부처 합동)¹⁸⁾

2013년 12월 관계부처 합동으로 수립한 빅데이터 산업 발전전략은 다음과 같은 배경으로 추진되었다. 첫째, 빅데이터가 ICT 분야의 새로운 패러다임이자 신성장 동력으로 급부상하고 있다. 둘째, 빅데이터 ‘자체’로뿐만 아니라 ‘활용’을 통한 경제사회 발전의 원동력이다. 셋째, 주요국 및 글로벌 기업은 빅데이터 ‘산업’ 육성 및 ‘활용’에 주력하고 있다. 하지만, 한국은 통신과 제조업 등 데이터 생산량이 많은 산업의 잠재력은 크지만 불확실성에 따른 투자리스크 때문에 아직 활용은 ‘저조’하다고 진단하고 있다. 그래서 국가차원의 빅데이터 ‘활용’을 촉진하고 빅데이터 ‘산업’을 육성하기 위한 전략적 지원이 필요한 시점임을 강조하고 있다.

〈표 Ⅲ-1〉 빅데이터 산업 발전전략(2013) 세부추진과제

세부 추진과제	기관	추진일정
1. 시장 창출 및 활용 촉진		
1-1 시범사업으로 Best Practice 조기 확산	미래부(안행부 등)	2013-2017
1-2 주요 산업별 선도 활용 프로젝트 추진	미래부(산업부, 안행부 등)	2014-2017
1-3 활용 인프라 제공 및 사회적 붐 조성	미래부(중기청, 안행부 등)	2013-2017
2. 산업 육성기반 확충		
2-1 핵심 원천기술 조기확보 및 국제 표준화 주도	미래부(산업부 등)	2014-2017
2-2 데이터 전문인력 양성 및 일자리 연계	미래부(교육부, 고용부 등)	2013-2017
2-3 거버넌스 정비	미래부	2013-2017
3. 지속발전가능 생태계 조성		
3-1 사용자 친화적 데이터 개방 확대 및 유통 촉진	미래부(안행부 등)	2013-2017
3-2 중소·중견기업 성장 촉진	미래부(금융위, 중기청 등)	2013-2017

출처: 관계부처 합동, 창조경제 및 정부3.0 지원을 위한 빅데이터 산업 발전전략 16쪽

18) 2013.12.11, 관계부처 합동, 창조경제 및 정부3.0 지원을 위한 빅데이터 산업 발전 전략

빅데이터 산업 발전전략으로는 빅데이터 산업의 인프라 측면에서 지속발전 가능한 생태계를 조성하고, 시장창출 및 확대라는 수요의 측면, 산업 육성 기반 확대라는 공급의 측면을 강조하고 있다. 이를 바탕으로 2배 이상의 시장창출 지원, 7개 분야 핵심기술 개발, 5천 명 이상의 고급인력 양성, 글로벌 전문 기업 10개 육성을 목표로 하고 있다. 주요 추진과제는 수요 측면에서 시범사업 등으로 Best Practice 조기 확산, 주요산업별 선도 활용 프로젝트 등, 활용 인프라 제공 및 사회적 붐 조성이 있다. 공급 측면에서는 산업 육성 기반 확충, 데이터 전문인력 양성 및 일자리 연계, 거버넌스 정비, 인프라 측면에서는 사용자 친화적 데이터 개방 확대 및 유통 촉진, 중소·중견기업 성장 촉진이 있다. 세부 추진과제별 추진기관은 <표 III-1>과 같다.

나. 주요부처의 빅데이터 활용계획

1) 통계청

통계청은 2015년 10월 빅데이터를 활용한 통계작성을 위해 통계데이터 허브국에 빅데이터통계과를 신설해 운영 중에 있다. 빅데이터 산업 정책과 활용에서 역할을 찾던 통계청은 빅데이터를 활용한 통계작성을 위한 제도적 기반을 구축하고 데이터 간 융·복합 사례를 발굴해 빅데이터 활용 확산에 나서고 있다. 통계자료와 빅데이터의 연계와 융합에 가장 큰 제도적 걸림돌은 개인정보 보호라는 측면에서 개인정보를 보호하면서 데이터 간 연계를 통해 빅데이터의 잠재가치와 활용에 공감대를 만들어간다는 기본적 방향을 갖고 있다. 기본적으로 통계목적으로 수집된 개인정보는 공유가 가능하지만 개별법에 의해 공유가 금지되어 있어 현재는 연계와 분석에 한계가 있기 때문이다. 그래서 통계청은 통계청의 기본 역할인 통계를 목적으로 하는 데이터를 수집을 근거로 데이터를 연계하고 분석결과를 확산하는 역할을 하는 데이터 허브 기능을 수행하고자 하고 있다. 이 과정에서 데이터 간 연계에서 발생하는 개인정보 보호에 대한 우려를 불식시키기 위해 개인정보 비식별화 방식을 검토하고 실제 적용한 연계사례를 확산하는데 노력을 기울이고 있다.

데이터 간 연계라는 측면에서 통계청에서 추진하고 있는 행정자료 통계

작성을 위한 부처 간 협업체계 강화와 활용도 제고를 위한 표준화, 행정 DB(통계등록부) 연계 및 통합을 추진하는 것은 큰 의미가 있다. 통계등록부는 행정자료를 이용해 작성한 명부자료로 조사통계의 모집단과 같은 기능을 수행한다. 행정자료가 표준화되고 연계가 가능해지면 파편화된 행정자료에서 연계된 데이터의 역할을 하게 되며 이 자체로 유용한 빅데이터가 되는 것은 물론이고 민관자료와 연계할 때에도 활용도가 높아져 빅데이터의 활용과 확산에도 매우 중요하다고 할 수 있다.

예를 들어 인구/가구 정보, 사업체/기업체 정보, 주택/건물 정보, 경제활동 정보에 대한 통계등록부 구축 방법은 주택과 건물 DB의 주소 정보를 활용해 인구/가구 정보 및 사업체와 기업체 정보와 연계가 가능하다. 연계된 개별 정보는 개인식별 정보로 개인의 경제활동 정보와 연계가 가능하고 개별 사업자 정보를 통해서는 사업체 조사의 정보와 연계가 가능하다. 이렇게 연계된 통계등록부는 개별 조사를 위해 수집한 정보를 연계해 새로운 통계를 생산할 수 있는 기반을 마련하게 된다.

통계청은 개별 법령의 제약으로 수집하고 있지 못한 행정자료 수집의 근거를 마련하고 개별법상 자료제공대상을 제한하고 있는 경우에도 통계 작성을 위한 목적으로 공유하기 때문에 이와 관련된 법을 개정하거나 정비 한다는 계획이다.

2) 국민안전처

국민안전처는 2016년 4월 전통적인 재난·안전 업무방식에 빅데이터를 접목 시키기 위해 재난예방 및 대응, 정책수립 지원, 활용기반 구축 등을 목표로 빅데이터 활용 종합계획을 수립했다. 계획에는 2020년까지 5년 동안 12개의 신규과제와 5개의 기존과제를 포함 총 17개 과제가 포함되어있다. 추진 과제별 주요내용을 살펴보면 <표 III-2>와 같다.

새롭게 발굴된 재난예방 및 대응 과제를 통해 사전 안전사고 예보를 통한 인명피해 저감, 피해자 수 예측을 통한 인력 배치 및 구조 효율과, 수사 조직·인력 및 장비의 적정 운영으로 해상 범죄율 감소, 화재위험도가 높은 시설물 사전 예방 강화 효과를 기대하고 있다. 정책지원의 측면에서는 생활

26 ●●● 여성가족 빅데이터 활용을 위한 기초연구

안전에 관한 정책 수립 시 과학적 타당성 지원, 맞춤형 구조서비스 구현, 학교폭력 예방을 위한 중점관리 추진, 자살징후 분석을 통한 자살 사망자 수 감축을 기대하고 있다. 활용기반 측면에서는 인력 확보, 법적, 제도적 기반 마련, 데이터 활용을 위한 표준화, 인프라 확보 등을 기대하고 있다.

〈표 Ⅲ-2〉 국민안전처 빅데이터 활용 신규과제 주요 내용(2016)

과제 제목	주요내용
1. 재난예방 및 대응	
1-1 주간 안전사고 예보 고도화	기존 주간 안전사고예보의 위험도 산정방식 개선 및 위험예보 모델 고도화(폭염, 한파 등)
1-2 모바일 공간정보 활용 재난 피해자 밀집도 파악	모바일 이동통신사의 기지국 정보를 활용, 모바일 공간통계지도 개발 및 재난발생 직전의 피해자 수 예측
1-3 해상범죄 발생 통계 데이터를 활용한 선제적 대응	해역별 불법어업, 실종변사, 밀입국 등 범죄유형과 범죄자의 수법 등을 패턴 분석하여 해경장비현황을 활용한 해역별 취약요소 선정
1-4 빅데이터 활용, 화재위험도 예측시스템 구축	건물 화재 위험요인별 종합정밀점검 및 기능 작동 점검결과 등의 빅데이터 분석을 통해 특별조사가 필요한 대상 선정 및 현장 정보 지원
2. 정책지원	
2-1 생활안전 민원 자료 분석	국민권익위원회 및 다산콜센터 등 국내에서 발생한 민원자료 등을 수집 및 분석하여 생활안전 분야 업무 추진 시 활용할 수 있는 위험인자 도출 및 활용안 마련
2-2 119구조 활동 분석	한정된 구조인력의 효율적 운영을 위해 구조 활동 정보(구조대 위치별 사고유형, 출동 및 현장 활동 소요시간) 등을 종합적으로 분석 추진
2-3 학교폭력 위험지수 개발	학교폭력 발생통계와 영향인자간의 연관성 분석, 지역별 학교폭력 발생 위험도 모델 개발 등
2-4 지역안전지수 분석(자살분야)	자살에 영향을 주는 주요인자 및 통계분석 자살 징후분석을 위한 자살위험도 모형 개발 및 지역 안전지수 자살분야 개선안 마련
3. 활용기반구축	
3-1 빅데이터 분석·활용 전문인력 양성	대학 및 민간 전문교육기관을 통한 빅데이터 교육 참여지원 및 학습 동아리, 포럼 등을 통한 활용환경 조성
3-1 재난발생시 개인정보 활용을 위한 법제도 개선	빅데이터 활용 시 개인정보보호 등에 대비하기 위한 제도개선 추진 (재난안전법에서 개인 정보 활용내용 추가방안 검토)
3-3 빅데이터 활용 단계별 관리 체계 구축	재난안전 데이터 값의 오류, 표준화 미비 등 데이터의 생산 및 수집 단계에서의 체계적 운영방안
3-4 재난 빅데이터 분석 장비구축	재난안전 과학기술 고도화를 위한 하둡 기반의 빅데이터 분석 시스템 구축으로 재난안전 빅데이터 허브기능 수행

출처: 국민안전처, ‘빅데이터 활용 종합계획 수립’ 보도자료 3쪽

다. 정부 주도의 공공 빅데이터 활용 사례

1) 2014 공공 빅데이터 분석사업

2014년 정부(총괄: 미래창조과학부, 주관: 정보화진흥원)가 데이터 분석 기반 미래전략 수립에 활용하겠다는 의지를 표명한 부처의 과제 공모를 통해 최종 4개 부처(기상청/농촌진흥청, 여성가족부, 특허청, 관세청)를 선정해 컨설팅을 진행했다.

〈표 Ⅲ-3〉 기관별 분석자료 형태 및 주요 분석내용

기관	자료 형태	분석내용
기상청	기상청 기상관측정보(공공)	기상기후와 농작물 생산량 관계 분석
	통계청 생산량 정보(공공)	병행충 발생 관련 비정형데이터 텍스트 마이닝 분석
	웹크롤링 비정형데이터(병해충)(민간)	
특허청	웹크롤링 비정형 데이터(민간)	기술의 시장수용성 분석
		기술의 시장실현성 분석
		기술의 시장선도성 분석
		기술의 시장융합성 분석
여성가족부	'13, '14 새일센터 데이터(공공)	구인기업 특성 분석
		구직자 특성 분석
		취업자 특성 분석
		미스매치 원인분석
관세청	5년간 관세청 수입 데이터(공공)	수입 통관실적 분석
	웹크롤링 비정형데이터(민간)	전자상거래 관심품목 후보 도출
		관심 품목 후보별 특성 분석

- 기상청/농촌진흥청 : 기상기후 데이터 분석을 통한 과학적 농업경영 지원
 - 농작물 생산에 영향을 미치는 기상요인 분석을 통해 ‘농작물 생산량

28 ●● 여성가족 빅데이터 활용을 위한 기초연구

예측 모형' 개발 및 국민경제 안정화 지원

- 연관기관 : 한국농수산식품유통공사, 한국농촌경제연구원
- 특허청 : 특허기술정보를 활용한 미래 ICT 유망 기술 예측
 - 특허정보 분석지표와 비정형 빅데이터를 매쉬업 분석해 국가특허전략 청사진 사업 부품분야 기술의 미래 유망성 예측
 - 연관기관 : 한국지식재산전략원
- 여성가족부 : 데이터 기반의 경력단절여성 맞춤형 재취업 지원
 - 새일센터 데이터의 다차원 빅데이터 분석을 통해 경력단절여성의 일자리 미스매치 원인을 진단하고 원인별 맞춤형 취업 지원
 - 연관기관 : 한국여성정책연구원
- 관세청 : 전자상거래 수출입 통관 데이터 분석을 통한 대국민 정보 제공
 - “전자상거래 유망 시그널 103選” 백서를 통해 건전한 시장 경쟁 유도 및 신규 일자리 창출을 위한 시그널 상품 정보 제공
 - 연관기관 : 한국관세무역개발원

데이터 기반의 미래전략 컨설팅 과제 중 여성가족부의 컨설팅 결과를 구체적으로 살펴보면 여성가족부 e-새일시스템의 데이터 총 110여만 건 중 데이터 전처리 과정 중 데이터 품질 저하 문제로 2012년 데이터를 제외하고 2013년과 2014년 상반기 5개 Data set의 약 67만 건의 데이터를 수집해 분석에 활용했다. 분석에 활용한 e-새일 시스템의 주요 내용은 <표 III-4>와 같다. 기 수행된 분석에서 추가된 문항은 e-새일센터에서 기 보유하고 있는 정보를 분석이 가능한 형태로 전처리 과정을 거친 것이다. 본 컨설팅을 통한 분석에서 가장 중요한 것은 e-새일센터 항목과 주요 데이터 간의 연계 분석을 시도했다는 점이다.

〈표 Ⅲ-4〉 e-새일센터 분석 수행 내용

공통	구인	구직/취업
직종별	4대보험 제공별	연령별
학력별	복리후생제공별	경력단절기간/사유별
경력조건별	휴일 방식별	구직등록 목적별
고용형태별	근무예정지역	찾아가는 취업지원 여부별
임금지급형태별	구인등록 목적별	결혼여부별
취약계층별		희망 근무지역
직무능력별		직업교육훈련
전공별		
외국어 능력/수준		

출처: 한국정보화진흥원(2014), 데이터 기반의 미래전략 컨설팅 완료보고서 - 여성가족부 내용 재구성

주요 분석별 데이터의 구체적 연계사례로 구인기업 특성을 분석하기 위해 구인기업 데이터와 통계청의 전국 사업체 조사 자료를 연계했고, 구직자와 취업자 특성을 파악하기 위해 통계청의 국내 여성고용 현황 자료(지역별 고용조사, 경력단절 여성 등의 경제활동 실태조사 등)를 연계했다. 또한, 개별 연계에서 한 발짝 더 나아가 일자리 미스매치 GAP 분석을 위해 구인기업 데이터, 구직자 및 취업자 데이터를 활용한 분석을 통해 경력단절 여성의 일자리 미스매치 원인을 파악하고자 하였다.

연계분석 과정을 통해 도출된 향후 발전방향을 살펴보면 빅데이터를 활용하는 과정에서 발생하는 데이터 간의 연계 문제와 표준화 문제를 지적하고 있다. 데이터 수기 입력에 따른 미입력 항목의 발생, 논리적 오류 데이터가 발생의 문제와 워크넷 시스템을 활용한 데이터 입력 방식으로 제약이 발생하게 되며 이로 인해 데이터 품질 저하로 주요 변수의 분석이 분석대상에서 제외되는 문제 등이 발생한다.

2) 국민건강보험공단 : 국민 건강·질병표준DB 구축사업¹⁹⁾

국민건강보험공단에서 추진한 본 사업은 지역단위를 대표하는 만성질환 지표에 대한 요구가 증가하고 사업장별 특성을 고려한 효율적인 근로자 건강관리 필요성의 대두, 생활환경 건강위험 요인관련 데이터 현황조사 필요성 증가가 본 사업의 배경이다.

본 사업을 통해 지표 산출을 근거로 한 효과적인 정책집행 및 건강수준의 증대, 실증적 데이터 기반으로 지역·사업장별 건강관리 정책 수립, 데이터 근거의 예산절감 및 산업 생산성 증대를 목적으로 한다.

○ 기초 데이터

- 건강보험 빅데이터 : 국민건강보험법 및 노인장기요양보험법에 의한 자격관리, 4대 사회보험의 통합징수, 요양급여비용 지급 및 사후관리, 건강검진, 장기요양관리 업무 등 업무수행과정에서 축적된 업무처리 시스템 자료를 통해 재구축된 빅데이터(국민건강정보DB)

〈표 Ⅲ-5〉 국민 건강·질병표준DB 지표 종류

구분		지표 종류(공공)
건강검진 수검률 지표		1차일반 생애검진수검률, 2차일반 생애검진수검률, 암검진통합 수검률, 위암검진수검률, 대장암검진수검률, 유방암검진수검률, 자궁경부암검진수검률, 간암검진수검률
건강위험요인 지표		흡연율, 음주율, 신체활동률, 비만율, 고도비만율, 초고도비만율, 복부비만율, 대사증후군 발견율
만성 질환 의료 이용 지표	고혈압	고혈압 의료이용률, 고혈압 신규의료이용률
	당뇨병	당뇨병 의료이용률, 당뇨병 신규의료이용률
	이상지질혈증	이상지질혈증 의료이용률, 이상지질혈증 신규의료이용률
	고혈압·당뇨병	고·당 의료이용률, 고·당 신규의료이용률
	고혈압·당뇨병·이상지질혈증	고·당·지 의료이용률, 고·당·지 신규의료이용률
	허혈성심질환	허혈성심질환 전체의료이용률, 허혈성심질환 입원의료이용률

19) 한국정보화진흥원(2014), 2014년 공공 빅데이터 분석사업 최종 결과물 - 국민건강 보험공단

Ⅲ. 빅데이터 정책의 여성가족 관련성 분석 ●● 31

구분		지표 종류(공공)
	뇌혈관질환	뇌혈관질환 전체의료이용률, 뇌혈관질환 입원의료이용률
	만성신부전	만성신부전 전체의료이용률, 만성신부전 입원의료이용률
	COPD	COPD 전체의료이용률, COPD 입원의료이용률
	간경화증	간경화증 전체의료이용률, 간경화증 입원의료이용률
	골다공증	골다공증 전체의료이용률, 골다공증 입원의료이용률
만성 질환 관리 지표	질병치료 및 합병증 지표	1년후 고혈압 투약순응률, 1년후 고혈압 정기방문율, 1년후 고혈압 조절률, 2년후 고혈압 조절률, 1년후 당뇨병 투약순응률, 1년후 당뇨병 정기방문율, 1년후 당뇨병 조절률, 2년후 당뇨병 조절률, 1년후 이상지질혈증 투약순응률, 1년후 이상지질혈증 정기방문율, 1년후 이상지질혈증 조절률, 2년후 이상지질혈증 조절률, 1년후 고당 정기방문율, 1년후 고당지 정기방문율, 1년내 심뇌신 신규의료이용률, 2년내 심뇌신 신규의료이용률, 3년내 심뇌신 신규의료이용률 등

출처: 국민건강보험(2015), 지역단위 맞춤형 건강관리를 위한 의료이용지표 활용 매뉴얼 7쪽

○ 지표 산출범위

- 지역단위 지표 제공범위 : 연도별/가입자 구분/지표별
- 연도별 : '02~'14년
- 지역단위: 시·도, 시·군·구, 읍·면·동
- 성별(2그룹) : 남, 여
- 연령(6그룹) : 20대, 30대, 40대, 50대, 60대, 70대 이상
- 가입자 구분(3) : 직장, 지역, 의료급여
- 소득수준(6그룹) : 보험료 1-5분위, 기타 분위

3) 부산광역시 : 부산도시서비스분석 정보시스템 구축²⁰⁾

정부 3.0에 부합하는 개방, 공유, 소통, 협력으로 시민이 활용할 수 있는 도시서비스 분석시스템 구축이 목표이다.

- 부산서비스인구 정보 제공 : 부산서비스인구 및 기타자료 분석마트 구축, 사용자 편의성을 고려한 서비스인구정보 제공

20) 한국정보화진흥원(2014), 2014년 공공 빅데이터 분석사업 최종 결과물 - 부산광역시

32 ●●● 여성가족 빅데이터 활용을 위한 기초연구

- 부산 통계지도 서비스 제공 : 인구, 주택, 교통 등 다양한 분야의 통계표를 시각화, 구군별, 읍면동 단위의 통계자료 등 공표자료 제공, 최대한 모든 시계열 제공, 가능한 통계표의 원 제공기관에서 자동 수집토록 설계하고, 불가한 경우 통계청 DB에서 자동 연계
- 기초 데이터
 - 부산시청 : 방법 CCTV, 버스정류장, 사업체 정보, 관광지 정보(공공)
 - 국토 지리정보원 : 100m 격자기반 데이터(공공)
 - 경찰청 : 여성 안심구역(공공)
 - 공공데이터 : 지하철, 학교, 경찰서, 건물대장(공공)
 - SK 유동인구 빅데이터(통계청 승인): 성/연령별 유동인구, 유집지별 시간대 유동인구, 유입지별 성/연령 유동인구, 요일별 유동인구, 시간대별 유동인구(민간)

4) 인천광역시 : 인천광역시 내·외국인 유입인구 분석²¹⁾

실증적이고 과학적인 방법을 통해 이동통신사 빅데이터(SKT 지오비전) 기반 국민 이동행태 분석을 통해 국민이동 행태에 대한 실증적 데이터를 확보해 향후 관련 마케팅 방안을 모색하는데 그 목적이 있다.

- 기초데이터
 - 이동통신사 유동인구 빅데이터(SKT 지오비전) : 성별(2그룹), 연령대(6그룹), 10개구, 10개 지역(지점)(민간)

5) 통계청 : 공공·민간 빅데이터 연계 현황 및 활성화 방안 - 신혼부부 통계를 중심으로²²⁾

개별적인 빅데이터 활용에서 새로운 가치 창출을 위한 데이터 연계의

21) 한국정보화진흥원(2014), 2014년 공공 빅데이터 분석사업 최종 결과물 - 인천광역시
 22) 통계청(2016), 공공·민간 빅데이터 연계 현황 및 활성화 방안 - 신혼부부 통계를 중심으로 보도자료

Ⅲ. 빅데이터 정책의 여성가족 관련성 분석 ●●● 33

필요성에 따라 개인정보 침해 없이 공공데이터와 민간데이터 연계의 제도적 기반 구축과 공공데이터 개방 추진 사례로 개인식별정보의 비식별화 방법이 적용되었다.

○ 기초데이터

- 인구통태, 인종 및 등록센서트, 주택소유통계(통계청) : 혼인날짜, 연령, 직업, 학력, 출산자녀수, 주택형태, 가구 구조, 보유주택수 등(공공)
- 전년도 보수총액(국민건강보험공단) : 공공
- 신용정보회사의 부채·카드 사용액 등(민간) : 부채 및 연체금액, 대출 건수, 대출유형 및 상환방식, 신용 및 체크카드 사용액, 추정소득 등

○ 공공 데이터와 민간 빅데이터 연계 방식

- 통계청이 공공데이터를 비식별화하여, 민간 기관에 제공(민간은 개별 법상 데이터 제공 불가)
- 민간기관이 2개 데이터를 연계하고, 연계키는 삭제한 분석용 DB 구축, 통계청 공유(특정 개인에 대한 식별이 불가능한 데이터)
- 데이터 제공, 분석 등 모든 과정을 통계청 내 통계데이터센터에서 수행, 분석결과만 집계표 형태로 반출(영국, 미국 통계청 방식)

6) 통계청 : 농가판매가격조사 현장조사 폐지

농가판매가격조사 자료수집 방식을 표본 현장조사 방식에서 농협중앙회가 보유한 빅데이터를 활용하는 방식으로 전환하였다. 통계청은 농협중앙회와 자료공유에 대한 협의를 완료하고 농협 빅데이터 기반 통계작성을 위한 시험 연구를 진행했다. 통계청은 민간 및 공공부문이 보유한 빅데이터를 통계 작성에 활용하는 국가통계생산의 패러다임 혁신을 점차적으로 확대해 나갈 계획이다.

라. 공공부문 빅데이터 활용의 특징

1) 공공 빅데이터 활용 주체

개방, 공유, 소통, 협력이라는 정부 3.0 기조에 따라 모든 공공기관에서 공공데이터에 대한 개방이 본격화 되었다. 공공데이터 개방에 따라 공공부문 빅데이터에 대한 관심과 활용 수요가 증가했다.

현재 존재하는 공공데이터의 빅데이터 활용 주체는 개인, 공공, 민간으로 구분해 살펴볼 수 있다.

- 개인 : 공공데이터 활용 분석, 개발(앱, 서비스), 보고서·논문 작성 등
- 공공 : 시스템개발·DB구축 등 빅데이터 생산(대국민 서비스), 중앙정부, 지자체 정책 수립 및 평가, 정책연구 등
- 민간 : 개별 민간 기업의 사업, 서비스 제공 과정에서 습득한 개인 정보를 통한 컨설팅, 마케팅 활용, 빅데이터 분석·판매, 공공데이터를 활용한 비즈니스 모델 개발 등

2) 공공부문에서의 빅데이터 자료 활용 특징

공공부문의 빅데이터 활용이 증가함에 따라 자료 활용 유형을 살펴보면 공공자료 활용, 민간자료 활용(데이터 크롤링 포함), 공공-민간자료 연계(보완) 활용으로 구분할 수 있다.

- 공공자료 활용 : 기 생산된 자료 또는 연계한 자료를 분석에 활용하는 것으로 앞서 살펴본 국민건강보험공단의 국민 건강·질병표준DB 구축 사업이 해당함.
- 민간자료 활용 : 민간영역에서 보유한 빅데이터 뿐 아니라 데이터 크롤링을 통한 자료 수집을 통한 활용에 이르기까지 민간영역의 자료를 활용한 분석으로 민간 기업의 컨설팅, 마케팅 분석이 대표적. 공공영역에서 SNS, 포털 분석으로 대표되는 데이터 크롤링을 통한 빅데이터 분석이 있음.

- 공공-민간자료 연계(보완) : 통계청의 사례처럼 공공데이터를 제공해 민간자료와 연계한 후 다시 이를 재분석하는 과정을 거친 유형은 공공 데이터와 민간데이터를 연계한 대표적인 사례라고 할 수 있음. 앞서 살펴본 이동통신사의 유동인구 빅데이터와 공공데이터를 연계해 GIS 시스템을 구축한 사례도 이에 해당함. 이와는 별개로 공공 데이터 분석결과와 민간데이터 분석, 데이터크롤링 분석을 시행한 후 이를 종합 분석하는 방식은 각 영역의 빅데이터를 보완 활용한 것으로 구분이 필요함.

3) 빅데이터 자료의 특징 : 데이터 연계와 개인정보 보호

데이터는 개별적으로 존재할 때 보다 연계되어 있을 때 활용의 폭과 범위가 크게 확대된다. 데이터를 생산하는 개별기관 내부의 데이터 연계부터 각 기관 간의 데이터 연계와 같은 이슈가 필연적으로 부각될 수밖에 없다.

공공기관의 경우 공공데이터 개방과 함께 개인정보 보호가 주요한 이슈로 부각되고 있고, 공공데이터 개방과 개인정보 보호는 서로 다른 곳을 지향하고 있다고 볼 수도 있다.

개인정보 보호의 수집, 처리, 보유 절차가 엄격해 짐에 따라 최소한 필요한 개인정보를 제외하고 폐기하는 기관이 많아짐에 따라 보유하고 있는 데이터의 양이 감소하는 모습도 보이고 있다. 주민등록번호와 같이 개인이 식별 가능한 정보의 경우 보유할 필요가 없는 경우 폐기함에 따라 성별 식별 정보도 함께 삭제되는 사례가 있다. 공공데이터 개방과 개인정보 보호의 상호 모순 상황을 해결하기 위한 방안으로 2016년 7월 개인정보 비식별화 가이드라인이 제시되었다. 하지만 비식별화 수행 주체와 적정성 평가에 대한 고민은 여전히 필요한 상황이다.

개인정보 비식별화는 개인의 정체성이 정보들과 연계되는 것을 방지하기 위해 사용되는 체계로 공개된 데이터 또는 기업 등이 보유한 데이터에서 개인을 유추 또는 식별할 수 있는 정보들을 특수한 처리를 하여 식별을 불가능한 상태로 만드는 과정이라 정의한다. 국가승인통계 중 행정통계 기반의 데이터의 경우 관련 규정과 근거법(통계작성)에 따라 유지가 되겠지만

향후 항목 축소의 우려가 제기되는 것도 사실이다. 비승인통계의 경우 개인 정보 보호 이슈에 따라 성별구분 통계의 비율이 더욱 낮아질 것이라는 우려도 제기되고 있다. 특히, 개인정보 보호 이슈로 인해 빅데이터의 생산과 활용의 측면에서 민간이 공공영역보다 경쟁력을 갖게 될 것이라는 우려가 제기되고 있고 신용정보회사, 통신사, 카드사 등을 통한 빅데이터 활용은 더욱 확대되고 있다.

4) 여성가족정책 빅데이터 연구와 데이터 생산 및 연계

본원은 여성가족정책 전문 연구기관으로 빅데이터를 활용한 연구가 본격화 되어야 한다는 필요성은 제기되고 있으나 구체적인 데이터 식별과 연계의 어려움 등의 문제로 어려움을 겪고 있다. 여성가족정책 분야의 빅데이터 활용은 기존 데이터의 제약을 받을 수밖에 없다는 점에서 한계를 가지고 있지만 새로운 통계의 발굴 및 활용이라는 측면에서 기회이기도 하다.

앞서 공공부문의 빅데이터 활용 주체별 활용사례 구분을 살펴보면 기존에 생산되는 데이터 간 연계방안을 마련하고 시스템과 DB를 구축하는 것도 주요 활용 방법으로 다루고 있어 다음과 같은 연구과제를 검토할 필요가 있다.

2. 여성가족 빅데이터 연구 분야

가. 여성가족 빅데이터 연구 필요 분야

1) 여성가족정책 관련 빅데이터 식별 연구

여성가족부를 중심으로 한 여성가족정책 관련 조사통계와 행정통계를 식별하고 개별 산하기관 및 연구기관의 사업실행과정에서 시스템적으로 발생하는 정보와 현재 축적되고 있는 정보는 무엇인지 식별하는 연구가 필요하다. 데이터 식별을 통해 활용가능성, 필요성 등을 평가하고 활용가능성과 필요성이 높은 정보에 대해서는 우선순위를 부여해 체계적인 데이터 축적과

연계를 고려한 표준화가 함께 진행되어야 한다. 또한, 법적 제도적 한계 때문에 활용이 불가능한 정보의 경우 법제도 개정을 통한 수집 기반이 마련되어야 할 것이다.

2) 여성가족정책 관련 빅데이터 생산 및 연계 방안 연구

향후 장기적으로 빅데이터 연구가 필요한 여성가족정책 연구의 주제를 발굴하고 빅데이터 생산방안을 검토하는 연구의 필요성이 제기된다. 생산의 개념은 새로운 데이터의 발굴 뿐 아니라 기존 데이터 연계를 포함하는 포괄적인 의미로 개별 부처 내에서 생산하는 데이터의 연계에서부터 부처 간 데이터의 연계, 공공데이터와 민간데이터의 연계까지 범위를 확장해 생산방안에 대한 종합적인 검토가 진행되어야 할 것이다.

3) 여성가족정책 관련 국가통계 연계 및 발전방안 연구

여성가족정책 관련 국가승인통계 중 빅데이터와 연계하거나 대체가 가능한 조사통계에 대한 검토 및 발전방안을 수립하는 연구이다. 정부 정책의 의사결정에서 데이터의 중요성을 점차 증대하고 있고 데이터 기반 환경에서 통계의 중요성은 더욱 강조되고 있으나 조사통계의 경우 조사거부와 불응 등 응답률이 저하되는 문제로 인해 조사 환경이 악화되고 있다. 1인 가구의 증가와 가구원 수의 감소로 주간 방문 시 응답률이 저하되고 있고 프라이버시 보호가 강화되는 사회적 분위기 속에서 최초 대면을 위한 접근이 차단되는 사례가 매년 증가하고 있다. 이러한 사회적 분위기 속에서 여성가족정책과 관련된 조사통계의 어려움은 더욱 높아질 것이 자명해 보인다. 특히, 경력단절, 다문화, 폭력, 성매매 등 민감성이 높은 조사가 상대적으로 많은 여성가족부의 상황을 고려할 때 데이터 연계를 통해 조사를 대체하거나 응답부담을 줄여 응답률을 높여나갈 수 있는 방안에 대한 검토가 요구된다고 할 수 있다.

4) 정부 주도 빅데이터 사업에 대한 성인지적 연구

범정부 차원 또는 각 부처별 빅데이터 관련 세부과제를 통해 진행되는 사업의 경우 사업추진을 위해 활용하는 데이터의 성별구분 정보와 이를 통해 생산되는 데이터의 성별구분 식별 정보에 대한 검토가 필요하다. 조사통계의 경우 성별구분 정보의 경우 응답부담과 유효 샘플 규모의 최적화를 통한 비용절감의 측면에서 모든 조사에서 성별을 구분해 조사하고 그 결과를 성별로 구분해 제공하고 있지 못한 실정이다.

하지만, 빅데이터를 활용한 데이터 생산과 분석의 경우 기존에 축적되거나 생산되고 있는 데이터를 활용해 분석하거나 연계해 활용하고 있기 때문에 활용하는 정보에 개인의 성별구분 정보가 포함되어 있는 경우 굳이 분석 항목에서 제외해 분석할 필요가 없어진다. 다만, 사용한 데이터가 성별구분이 불가능한 경우 왜 불가능한지에 대한 정보를 제공한다면 향후 관련 데이터 수집에서 성별구분 정보를 수집할 수 있는 방법과 데이터 수집 개선 방안의 기초자료로 제공될 수 있다는 점에서 큰 의의를 갖는다.

또한, 아직 통계청에서는 빅데이터 방법론을 통해 생산된 통계에 대해 국가 승인통계로 관리할 계획이 없기 때문에 작성과 승인과정에서 통계법에 따라 성별분리 여부를 점검하는 제도적 관리대상에서 사각지대에 놓일 가능성이 매우 크다고 할 수 있다.

5) 여성가족정책 빅데이터 DB 구축 및 운영

앞서 살펴본 빅데이터의 식별과정과 생산과정에 대한 연구가 진행되고 실제 데이터가 생산되기 시작하면 관련 정보를 제공하는 시스템의 필요성이 자연스레 커질 수 밖에 없다. 특히, 빅데이터의 경우 활용의 측면에서 기존 통계시스템에서 제공하는 결과표 만으로는 활용도가 매우 떨어진다. 또한, 데이터를 직접 분석할 수 있는 원자료 형태로 제공하기에는 데이터의 크기가 매우 큰 특징을 갖는다. 그래서, 누구나 해당 빅데이터에서 활용하고자 하는 키워드를 선정하면 분석 가능한 형태의 데이터로 제공하는 시스템의 필요성이 높아진다. 예를 들어, 한국언론진흥재단에서는 주요 언론사의

Ⅲ. 빅데이터 정책의 여성가족 관련성 분석 ●●● 39

뉴스를 수집하는 시스템을 통해 형태소 분석, 개체명 분석, 네트워크 분석 등 기사를 분석하는 뉴스분석 시스템을 구축하고 이를 저장하는 뉴스 저장 시스템인 빅카인즈(Big Kinds) 시스템(<http://bigkinds.or.kr>)을 구축해 서비스 하고 있다. 분석된 결과를 단순히 제공하는 서비스에서 한 발짝 더 나아가 특정 뉴스에서 정보를 추출할 수 있도록 개인화 된 텍사노미 관리, 질의어 관리, 정보 추출 패턴 규칙 적용 기능을 제공하고 트렌드, 네트워크 분석 등 다양한 분석도구를 통해 그 결과를 활용할 수 있는 빅 카인즈 프로(Big Kinds-Pro) 서비스는 빅데이터 활용 시스템 구축의 좋은 예라 할 수 있다.



IV

해외의 여성가족 빅데이터 활용전략 및 사례

- | | |
|------------------------|----|
| 1. 해외의 여성가족 빅데이터 활용 전략 | 43 |
| 2. 여성가족 빅데이터 활용 사례 | 56 |



1. 해외의 여성가족 빅데이터 활용 전략

가. 빅데이터 활용 전략

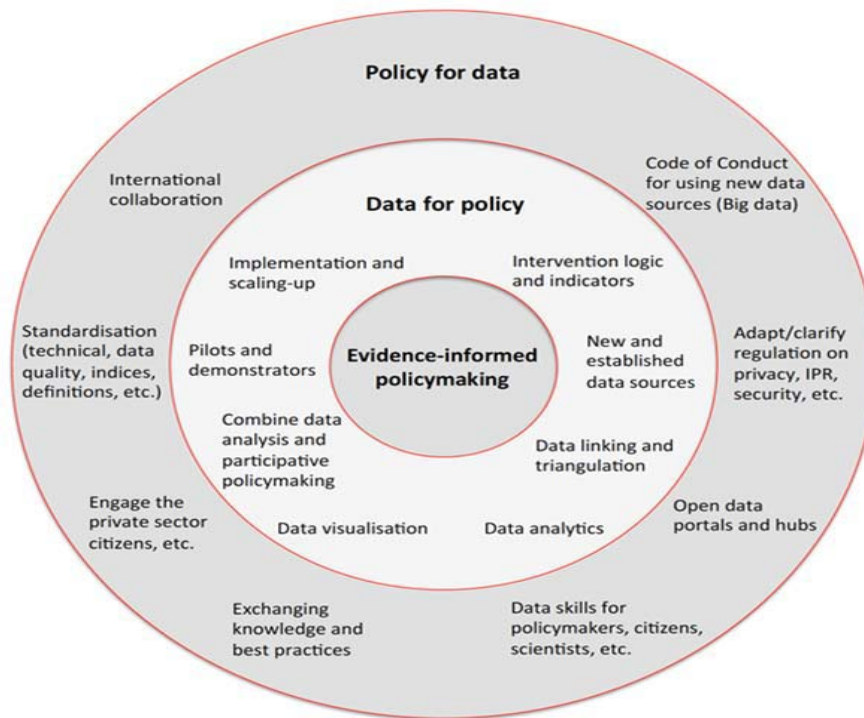
빅 데이터가 우리사회에 중요한 역할을 할 수 있을 것이라는 점에서 빅데이터의 중요성에 대한 인식이 증대되고 있다. 특히 경제학적인 측면에서 빅데이터의 가치는 중요하다. 일반 대중 및 사기업의 상업적 활용으로 인한 가치는 상당하다. 그러나 빅데이터의 새로운 지식 생산은 한 학문에만 국한된 것이 아니라 다른 학문에도 영향을 끼칠 수 있는 파급 효과도 존재하고 있다. 그러한 다학제적인 파급효과는 빅데이터가 공공정책과 일반적인 법적 정보 제공에 상당한 영향력을 끼칠 수 있다는 가능성이 존재할 수 있다 (LeFurgy, 2015:3-4).

대중이 정부의 데이터에 대한 관심과 활용은 중요한 시사점이 있다. 광범위한(broad-based) 의미에서의 대중 집단, 그러한 대중 집단들이 데이터를 활용하여 새롭게 데이터를 만들어 활용되어진 지식을 출판하여 신지식의 확산, 또 그로 인한 파급 효과는 결국 정부의 정책에 도움을 줄 수 있다. 그러나 대중이 정부의 자료에 대한 접근은 지식소유권에 대한 법적 침해측면에서 반대 의견이 존재한다. 이러한 반대 의견에도 관련 정책적 보완 및 가이드라인을 제시하여 빅데이터 형성 및 활용에 있어서 많은 대중들이 활용할 수 있게끔 할 수 있는 것이 중요하다(LeFurgy, 2015:8-16).

빅데이터 시대에 접어들어 선진 국가들은 세계를 선도하기 위한 글로벌 경쟁력을 갖추기 위하여 빅데이터를 이용한 정책개발과제 전략을 제시하고 있다. 미국에서는 “빅데이터 전략 수립 및 국가차원의 거버넌스 구축”을 통하여 국민들에게 필요한 빅데이터 정보 제공 및 국민들을 위한 보다 효율적인 정책 개발을 위해 노력하고 있다. 영국에서는 빅데이터를 활용하여 “정보공유·활용을 위한 데이터 개방중심의 정책”을 시행하고 있다. 싱가포르에서는 빅데이터를 활용하여 “국가 위기관리를 위한 정책지원 체계 마련”을 목표로 하고 있다²³⁾.

23) 스마트 국가 구현을 위한 빅데이터 마스터플랜. 2012. 교육과학기술부, 행정안전부, 지식경제부, 방송통신위원회, 국가과학기술위원회. 12-15쪽

빅데이터와 2030 지속적 발전을 위한 아젠다 관련 보고서의 빅데이터 정책 활용 전략을 그림으로 나타내면²⁴⁾ [그림 IV-1]과 같다. 실제적이고 기술적인 제한점에도 불구하고 기존 데이터와 비교하여 데이터의 상당히 큰 양, 데이터의 실제 상황 반영, 다양한 데이터 활용이 빅데이터의 장점으로 지적하고 있다. [그림 IV-1]에서 이러한 빅데이터의 장점을 활용하여 근거 중심의 정책 집행을 강조하고 있다. 기존 데이터를 수집하고 연결하여 새로운 데이터를 형성하거나 인터넷 및 SNS 상의 실시간 수집된 비정형 데이터를 정형화하여 정책 형성에 유용하게 사용할 수 있도록 하는 점이다(Abbas Maarooof, 2015).



출처: Abbas Maarooof, 2015. Big Data And The 2030 Agenda For Sustainable Development: Final Draft Report Open For Comment. 9쪽.

[그림 IV-1] 빅데이터와 정책의 관계

24) Abbas Maarooof, 2015. Big Data And The 2030 Agenda For Sustainable Development: Final Draft Report Open For Comment

IV. 해외의 여성가족 빅데이터 활용전략 및 사례 ●● 45

최근 미국 연방정부의 빅데이터 전략계획²⁵⁾을 살펴보면 아래와 같다.

첫 번째로, 다음 세대의 빅데이터 활용의 창의적 발전을 위해 지속적인 빅데이터 연구투자를 강조하고 있다. 지속적인 빅데이터 연구투자를 활성화 하여 다음 세대에게 지금 중요하게 등장하고 있는 빅데이터 활용 기술의 중요성을 전하며 미래의 세대에게 빅데이터 기술 향상이라는 책임감을 부여 할 전략을 제시하고 있다.

두 번째로, 빅데이터 연구결과의 신뢰성을 보장하기 위하여 적절한 연구 방법과 수리적 접근방법의 정확성을 강조하고 있다. 빅데이터로 형성된 데이터를 다른 연구자가 분석하여도 같은 연구결과를 가져올 수 있도록 하여 데이터의 연구결과의 불확실성을 최소화 한다. 빅데이터 연구결과의 신뢰성을 이해하고 탐색하여 빅데이터 연구결과로 혁신적 발견과 신뢰성 있는 행동 혹은 정책으로 이끌 수 있는 연구투자 전략을 제시하고 있다.

세 번째로, 사이버인터넷을 활용한 데이터 기반 구축의 중요성을 강조하고 있다. 실시간 축적되어질 수 있는 빅데이터는 과학, 국방, 상업 등에 이르는 다양한 분야에서 유용하게 활용할 수 있는 정보자원을 형성한다. 정보자원을 관리하고 분석할 수 있는 인적 자원이 필요하며 인적 자원에 대한 보다 나은 교육적 지원에 대한 전략이 제시됐다.

네 번째로, 빅데이터의 효용과 가치를 높이기 위하여 이용 가능한 데이터가 더욱 많아져야 하며 접근성이 용이하여야 한다. 빅데이터의 범위(scale)와 이질성(heterogenity)이라는 문제가 대두될 수 있지만 데이터 공유 및 관리를 할 수 있는 제도를 통하여 데이터 질을 증대시키는 전략이 제시됐다.

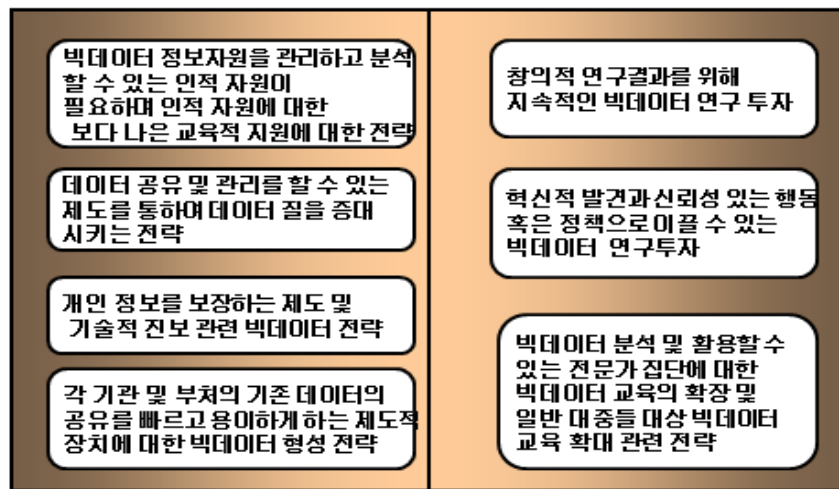
다섯 번째로, 빅데이터 자료 수집, 공유 및 활용 측면에서 개인 정보 보장 및 윤리적인 측면을 강조하고 있다. 개인 정보를 보장하는 제도 및 기술적 진보 관련 빅데이터 전략이 제시됐다.

여섯 번째로, 빅데이터 중요성이 인지되고 필요성이 증대됨에 따른 종합적 빅데이터 교육의 중요성이 강조되었다. 빅데이터 분석 및 활용할 수 있는 전문가 집단에 대한 빅데이터 교육의 확장 및 일반 대중들 대상 빅데이터

25) The Federal Big Data Research And Development Strategic Plan. 2016. The Networking And Information Technology Research And Development Program.

교육 확대 관련 전략이 제시됐다.

일곱 번째로, 기관 및 부처 간의 데이터 자료 공유에 대한 장벽을 없애는 점을 강조한다. 특히 국가 우선 정책에 해당되는 경우 각 기관 및 부처의 기존 데이터의 공유를 빠르고 용이하게 하는 제도적 장치에 대한 빅데이터 형성 전략이 제시됐다.



출처: The Federal Big Data Research And Development Strategic Plan. 2016. The Networking And Information Technology Research And Development Program. 재구성

[그림 IV-2] 미국 연방정부의 빅데이터 전략계획

나. 여성가족 정책 빅데이터 활용 전략

선진국의 빅데이터 전략 방향 사례가 우리에게 주는 시사점은 정부 차원에서 적극적으로 빅데이터 추진체계를 마련하고 정부 각 부처가 확보하고 있는 데이터를 연계하여 국민이 쉽게 활용할 수 있게끔 하고 정부 또한 빅데이터를 활용하여 경제, 안전, 가족, 보건 등 다양한 영역에 보다 효율적이고 효과적인 정책을 만드는 것이 필요하다는 점이다.

보다 효율적이고 효과적인 정책을 위해 사회 전반의 모든 정책에 성인지적인 정책이 강조되고 있어 여성을 위한 보다 나은 정책 개발을 위하여 빅데이터 활용이 필요하다. 여성가족 정책 형성 과정에서 빅데이터 활용은 두 가지

IV. 해외의 여성가족 빅데이터 활용전략 및 사례 ●●● 47

측면에서 설명 될 수 있다(Buvinic, M., Furst-Nichols, R., & Koolwal, G, 2014).²⁶⁾

첫 번째로 정부가 공공 데이터를 연계하고 수집하여 여성들이 쉽게 활용할 수 있게 하고 또한 정부가 공공 데이터를 모은 빅데이터를 활용하여 여성을 위한 보다 나은 정책을 수립하는 것이다. 두 번째로, 소셜 네트워크를 활용한 빅데이터 구축 및 활용이다. 빅데이터를 활용하여 여성의 행동 및 정책에 대한 여성의 의견 및 견해와 입장을 이해하고 여성 정책 수립 및 평가 시 여성의 의견 및 견해를 고려하는 것이다.

여성가족 정책과 직접적으로 관련된 충분한 데이터가 결국 빅데이터로 연계될 수 있음에도 불구하고 여성과 관련된 데이터가 부족하다는 점이다. 빅데이터 활용으로 잠재적으로 여성가족 정책에 기여할 수 있지만 실제 활용할 수 있는 데이터를 여성들이 쉽게 활용 가능하고 또한 공무원들이 빅 데이터를 활용하여 여성가족 정책을 만드는 일은 결코 쉽지 않다는 점이다.

주요한 이유로 빅데이터를 형성할 기존 데이터의 문제점이 지적되고 있다 (Buvinic, M., Furst-Nichols, R., & Koolwal, G, 2014). 건강, 교육, 경제활동 기회, 정치 참여, 안전에 관한 여성가족 정책의 기존 데이터가 부족하다는 점이 지적되고 있다. 이는 여성가족 정책의 빅데이터 활용에 대한 근본적인 문제점을 제시하고 선진국 빅데이터 전략 방향에 대한 출발점이 됨을 시사하고 있다.

첫 번째로, 건강 분야 측면에서 여성건강에 대한 이해는 모성건강에 국한된 데이터 뿐 만이 아니라 여성의 생애주기 다발성 질환 및 여성 특이적인 질환에 관한 데이터가 필요하다. 또한 건강에 대한 이해는 의료적이며 치료적인 측면에서 뿐만이 아니라 질환의 예방측면에서 생물학적 및 사회환경적 요인들을 동시에 고려한 데이터가 필요하다.

두 번째로, 교육 분야 측면에서 교육에 관한 성별 분리 통계를 통한 교육 데이터의 사회적인 파급 효과는 상당히 크다. 교육 취약 계층으로서 여성에 대한 정보가 더욱 많이 필요하며 관련 데이터가 필요하다. 또한 국제 비교

26) Buvinic, M., Furst-Nichols, R., & Koolwal, G. (2014). Mapping gender data gaps, UN Foundation. Retrieved August 30 from http://data2x.org/wp-content/uploads/2014/11/Data2X_MappingGenderDataGaps_FullReport.pdf

연구를 통한 성별 교육의 효과를 측정할 수 있는 질 좋은 데이터의 부족함을 지적하고 있다.

세 번째로, 경제 분야 측면에서 여성일자리 창출이 가장 중요한 이슈 중 하나다. 여성 일자리 현황 파악에서 여성이 남성 보다 경제활동으로 참여하고 있는 비정규직에 관한 정확하고 신뢰성 있는 데이터가 더욱 더 필요하다. 또한 비정규직에 근무하는 여성들에 대한 자세한 정보를 가질 수 있는 데이터가 필요하다. 경제활동으로 인한 수입 측면에서 남성과의 차이, 농업에 종사하는 여성 경제 활동 등 좀 더 자세한 정보를 가진 데이터가 필요하다. 특히 모바일 폰 혹은 인터넷 등의 영역에서 여성의 삶에 미치는 영향력(동료와의 정보 소통, 구직 활동, 소비자 물품의 가격 정보, 금융 거래, 신기술 획득 등)에 관한 데이터가 필요하다고 지적하고 있다.

네 번째로, 여성 정치 참여에 기여할 수 있는 데이터의 부족을 지적하고 있다. 남성 보다 취약 할 수 있는 여성들의 정치 참여에 대한 데이터가 필요하며 지역수준 혹은 풀뿌리 정치참여 수준에서 여성정치 참여 및 리더쉽 관련 데이터가 우선적으로 필요하다고 지적하고 있다.

다섯 번째로, 여성 안전에 관한 데이터가 상당히 부족함을 지적하고 있다. 특히 여성이 평화와 안전 활동 참여 및 리더쉽 관련 데이터가 필요하다고 지적하고 있다.

다. Data2X

Data2X는 빅데이터 분석을 기반으로 전세계 여성과 소녀들에게 보다 나은 삶을 제공하기 위하여 2012년에 형성된 협력적 플랫폼으로 United Nations (UN)의 우선계획(Initiative)로 수행되고 있으며 William & Flora Hewlett 재단, Bill & Melinda Gates 재단 및 Clinton 재단이 지원하고 있다. Data2X이 빅데이터 분석의 기준은 UN이 제시한 지속가능한 발전 목표(Sustainable Development Goals, SDGs)에 대한 성별영향평가이다. 이를 위하여 UN 산하 기관, 정부, 시민단체, 학계 등과 협력하여 젠더 데이터를 구축, 확장, 분석하여 효과적인 정책과 전략을 제공하고 의사결정을 제시한다.

Data2X는 UN의 17개 SDG 중 8개 목적에 대해서 소녀와 여성을 대상

IV. 해외의 여성가족 빅데이터 활용전략 및 사례 ●● 49

으로 측정할 수 있는 지표를 다음과 같이 확인하였다²⁷⁾. 8개 목적은 빈곤(no poverty), 기아(zero hunger), 건강과 웰빙(good health and well-being), 교육의 질(quality education), 성평등(gender equality), 직업과 경제성장(decent work and economic growth), 불평등(reduced inequalities) 및 평화, 정의와 강력한 기관(peace, justice and strong institutions)이다. 이 중 세 번째 목적은 건강과 웰빙으로 여성건강과 관련된 것으로 이를 평가하기 위해서 다섯 가지의 분석 가능한 빅데이터 지표를 다음과 같이 선정하였다.

빅데이터로 분석 가능할 것으로 제안된 건강관련 젠더지표

목적3: 모든 연령에서 건강한 삶을 보장하고 웰빙을 증진시킨다.

- 출생 100,000명당 산모사망
- 5세 이하 여아 사망률과 5세 이하 여아와 남아 사망비
- 취약인구 1,000명당 여성 HIV 신생감염률과 남성과 여성 HIV 신생감염률
- 15-19세 연령그룹에서 여성 1,000명당 출산율
- 가임기(15-49세) 여성 중 현대 피임방법을 사용하는 여성의 백분율

Data2X는 빅데이터 분석을 위하여 필요한 자료를 확인하면서 성별 데이터 차이(gender data gap) 변수를 확인하여 이를 건강, 교육, 경제기회, 정치참여, 안전의 다섯 가지 영역에서 28가지 변수로 분류하였다.²⁸⁾ 분류한 변수에 대한 성별 데이터 차이는 다음의 4가지로 구분하여 제안하였다. 국가 자료의 부족 혹은 정기적 자료 산출의 부족, 비교할 수 있는 국제기준의 부족, 영역 간 정보의 부족으로 인한 복잡성 부족 및 구성변수를 비교하기 위한 데이터 세트의 구체성 미약. 구체적인 성별 데이터 차이가 나는 정도는 다음의 <표 IV-1>과 같다.

27) Data2X. (2014). Ready to Measure Progress on Gender Equality. Retrieved from <http://data2x.org/what-is-gender-data/ready-to-measure/>

28) Data2X. (2014). Gender data gaps. Retrieved from <http://data2x.org/what-is-gender-data/gender-data-gaps/>

50 ●●● 여성가족 빅데이터 활용을 위한 기초연구

〈표 IV-1〉 Data2X에서 제시한 성별 데이터 차이

영역	변수	성별 데이터 차이			
		국가자료 부족 혹은 정기적 자료 산출 부족	비교할 수 있는 국제적 기준부족	영역간 정보부족으로 복잡성 부족	구성변수비교 위한 데이터 세트의 구체성 미약
건강	모성의 질병 이환률과 사망률	○			○
	여성의 과도한 질병부담	○			
	여성에 대한 폭력	○		○	
	정신건강	○		○	○
	청소년/청소년 건강	○		○	○
	여성의 의료시스템 활용	○			○
교육	학습 성과	○	○		
	제외된 소녀	○	○		
	진학률	○			
경제 기획	무급 노동	○			
	비공식 고용	○			
	유급노동에 대한 수입과 기회비용	○	○		
	이민노동자 상황	○	○		
	취업이동	○	○	○	○
	기업가 정신	○		○	○
	자산 소유권	○		○	○
	농업생산성	○	○	○	○
	금융서비스 접근성	○			
	보육서비스 접근성	○	○	○	○
	휴대전화와 인터넷 접근성	○	○	○	○
정치 참여	지역지차, 정치단체 및 전문적 대표	○	○		
	국가 신분증명서 문서화	○			○
	투표 등록과 참가	○	○		
안전	질병 이환률, 사망률과 관련한 갈등	○			○
	강제 추방 혹은 이주	○			○
	젠더변수의 영향에 대한 갈등	○		○	○
	성적/젠더기반 폭력과 관련된 갈등	○		○	○
	평화와 안전에 관한 과정에 참여	○	○	○	○

IV. 해외의 여성가족 빅데이터 활용전략 및 사례 ●●● 51

이 외에도 Data2X는 빅데이터 분석을 통하여 전세계 여성건강을 도모하기 위한 소규모 프로젝트를 진행하고 있다(Data2X, 2015). 빅데이터를 이용한 성별분석 소규모 프로젝트는 휴대폰 이용 기록, 위성 데이터, 소셜미디어를 이용한 프로젝트가 있다. 먼저, 휴대폰 사용자의 성별과 사회경제적 지표를 파악하여 휴대폰 사용빈도, 충전 빈도, 통화패턴을 성별로 분석하여 의료기관과 정보서비스 이용에 관한 모델을 구축하여 의료서비스 이용을 분석하고 예측한다. 인공위성을 이용한 프로젝트는 Flowminder Foundation이 수행하고 Data2X가 지원하고 있으며, 여성의 복지, 사망률, 유병률, 영양부족, 피임율, 이동자유성 등에 관한 지리공간학적인 분석을 통하여 전세계적으로 여성의 건강에 집중해야 할 지역을 선정한다. 이 프로젝트에 포함된 국가에는 방글라데시, 부르키나 파소, 에디오피아, 아이티, 온두라스 및 네팔 등이 있다. 소셜 미디어 분석을 위한 프로젝트는 트위터를 이용한 빅데이터 분석모델을 구축하며 두 개의 프로젝트가 진행된다. 트위터에 포스트 된 우울관련 내용을 분석하여 남아프리카와 인도의 청소년들의 우울 유병률과 우울편중 지역을 확인하는 모델을 구축하기 위해서 UN Global Pulse가 파트너 기관으로 프로젝트를 수행하고 있으며, 청소년의 정신건강을 확인하기 위하여 긍정적, 부정적 트위터 메시지를 분석하는 프로젝트는 멕시코 통계와 지리에 관한 국가기관(National Institute of Statistics and Geography)와 협력하여 수행하고 있다.

〈표 IV-2〉 Data2X 수행 빅데이터와 젠더 프로젝트

빅데이터를 이용한 소규모 프로젝트 영역	건강관련 프로젝트 내용	협력기관
휴대전화 이용	여성의 의료기관 접근성 분석	통신사
인공위성 자료	여성 복지 및 건강관련 지표의 지역적 분석	Flowminder Foundation
트위터 자료	청소년의 우울유병률	UN Global Pulse
	청소년의 정신건강	Mexico National Institute of Statistics and Geography

UN은 빅데이터를 이용한 SDG의 성별평가를 위해 전세계적으로 다양한 기관과 협력체계를 구축하고 있다(표 IV-3 참조).²⁹⁾ 현재 앞에서 설명한 Data2X를 포함한 24개 기관과 공식적으로 협력체계를 구축하고 있으며 이 중 11개 기관이 SDG 5번째 목표인 성평등과 관련된 협력체계를 구축하고 있다. 앞서 설명한 Data2X를 제외한 성평등과 관련된 기관에 대한 간략한 설명은 다음과 같다.

Orange는 사적-공적 파트너십을 기반으로 모바일 네트워크에서 추출되는 자료를 이용한다. DataShift는 시민단체들에 의하여 생성된 데이터를 정책 개발에 활용한다. Humanitarian Data Exchange는 2014년에 시작되어 200개 이상 국가에서 4,000개의 데이터세트를 취합하여 데이터를 공유하여 빅데이터 플랫폼을 형성하고 있다. OPAL 프로젝트는 빅데이터를 이용하여 오픈 플랫폼과 알고리즘을 개발하는 것이다. Open Data for Development Network는 65개 이상의 기관과 글로벌 협력체계를 구축하고 있으며 저개발국가의 빅데이터 플랫폼을 구축하고 있다. 017은 정기적으로 공개된 빅데이터를 이용하여 분석을 하는 개인이나 단체들을 초대하여 SDG 목표에 대한 평가를 수행을 촉진시킨다. Sustainable Development Solutions Network는 SDG의 평가에 필요한 데이터를 명명하고 부족한 변수에 대한 데이터 생성을 목표로 한다. Ninga Effect은 여성의 스토리에 집중하여 수집된 자료를 비디오로 제작하고, 집에서 재배되는 작물의 활용법에 대한 인식을 개선시키는 등 기존의 자료를 활용하여 여성의 삶의 질을 증진시키고 있다. 이 프로젝트는 가나와 르완다에서 수행된다. Youth Action Mapper의 목적은 약 100개 국가에서 청소년 100만명의 주도하에 휴대전화를 이용한 SDG 관련 빅데이터를 수집하고 분석하는 것이다. 마지막으로 UNEP Live는 지식공유를 위하여 국가, 지역 및 세계의 데이터에 대한 플랫폼을 구축한다.

제시된 기관 중 한국의 자료를 포함하는 기관이 있는데 International Barcode of Life는 다양한 국가의 유전적 자료를 빅데이터를 이용하여 분석한다. 이 기관은 생물의 유전자를 지역별로 분석하여 생물학적 다양성을 빅데이터로 분석한다.

29) Global Partnership for Sustainable Development Data (2016). Data collaboratives. Retrieved from <http://www.data4sdgs.org/dc-data2x>

〈표 IV-3〉 Data2X 협력기관, 역할, 담당영역 및 담당국가

협력기관	기관역할								담당 SDG	담당국가
	새로운 자료 생성	현 시스템 개선	새로운 자료수집 기술 개발	기존 데이터의 개방, 활용, 공유	데이터 가용성과 요구역량 증진	이용가능한 자원의 활용	원칙과 수준 높은 중재의 개발 촉진	의사결정과 정책개발 지원	정부, 시민단체, 사적영역에 데이터 공유와 사용 증진	
AirNow				○		○		○	3,11,15,17	중국, 멕시코, 대만, 미국
DataActLab			○			○		○	1~17	가나, 나이지리아, 필리핀, 탄자니아
PEPFAR MCC		○		○	○		○	○	3,5,8,17	다르에스살람, 탄자니아
DATA for Climate Action				○	○	○	○	○	13,17	전세계 국가
Orange	○	○	○				○	○	1-5, 8-11, 17	코트디부아르, 세네갈
Data2X		○	○			○		○	5, 17	국가 명시안됨
DATASHIFT			○	○	○	○	○	○	5,16	아르헨티나, 케냐, 네팔, 탄자니아
GEOGLAM		○	○	○			○	○	1~3	국가 명시안됨
GeoSUR	○			○				○	1,3,9,11,13,15	국가 명시안됨
Governance Data Alliance			○	○	○		○	○	16	전세계 국가
Health Data Collaborative		○	○	○	○		○	○	3	국가 명시안됨

54 ●● 여성가족 빅데이터 활용을 위한 기초연구

협력기관	기관역할								담당 SDG	담당국가
	새로운 자료 생성	현 시스템 개선	새로운 자료수집 기술 개발	기존 데이터의 개방, 활용, 공유	데이터 이용성과 요구역량 증진	이용가능한 자원의 활용	원칙과 수준 높은 중재의 개발 촉진	의사결정과 정책개발 지원	정부, 시민단체, 사적영역에 데이터 공유와 사용 증진	
Humanitarian Data Exchange				○					1~9, 11, 13, 17	제네바, 나이로비, 뉴욕
International Barcode of Life	○			○		○	○		3, 9, 14, 15	아르헨티나, 호주, 브라질, 케냐, 중국, 콜롬비아, 코스타리카, 프랑스, 핀란드, 독일, 인도, 케냐, 마다가스카, 멕시코, 뉴질랜드, 네덜란드, 노르웨이, 파키스탄, 파나마, 라투아니아, 페루, 포르투갈, 러시아, 사우디아라비아, 남아프리카, 대한민국, 영국, 미국
JUDS Project				○		○		○	1~4, 10, 16	전세계 국가
MPPN	○	○		○				○	1	칠레, 콜롬비아, 코스타리카, 에콰도르, 엘살바도르, 말레이시아, 멕시코, 튀니지, 호치민 시티
OPAL						○		○	1, 3, 5, 8, 10, 11, 13, 16, 17	국가 명시 안됨

IV. 해외의 여성가족 빅데이터 활용전략 및 사례 ●● 55

협력기관	기관역할									담당 SDG	담당국가
	새로운 자료 생성	현 시스템 개선	새로운 자료수집 기술 개발	기존 데이터의 개발, 활용, 공유	데이터 이용성과 요구역량 증진	이용가능한 자료의 활용	원칙과 수준 높은 중재의 개발 촉진	의사결정과 정책개발 지원	정부, 시민단체, 사적영역에 데이터 공유와 사용 증진		
Open Data for Development Network	○		○	○		○	○	○		3~5, 8~9, 11, 16~17	전세계 국가
017				○	○				○	1~17	콜롬비아, 안도, 리투아니아, 나이지리아, 필리핀, 스위스, 탄자니아, 영국
Project 8				○	○	○		○		2	전세계 국가
Sustainable Development Solutions Network				○		○		○		5, 8~9, 11~12	보고타, 콜롬비아
Nzinga Effect								○		5	가나, 르완다
Youth Action Mapper					○				○	1~17	라이베리아, 나이지리아
UNEP Live					○	○			○	1~17	전세계 국가
USA Sustainable Cities Initiative				○	○				○	11	미국의 벨티모어, 뉴욕, 산호세

출처: Global Partnership for Sustainable Development Data (2016). Data collaboratives. Retrieved from <http://www.data4sdgs.org/dc-data2x>

2. 여성가족 빅데이터 활용 사례

가. 경제부문 활용사례

최근 UN 빅데이터 글로벌 워킹 그룹³⁰⁾에서는 상당히 많은 양의 데이터 사업을 소개하고 있다. 빅데이터 사업들을 소개하는 내용을 검토하고 성인지적 여부와 성별 분리 여부를 기준으로 분석하였다³¹⁾. 분석 방법은 사업의 주제와 사업소개 내용을 검토하는 과정을 통하여 경제영역 부문과 비경제 부문의 영역으로 구분하였다. 또한 빅데이터 사업 내용 중 ‘gender’와 ‘sex’ 혹은 ‘female’이라는 단어가 있는지, 내용이 여성이슈를 반영한 것인지를 기준으로 선정하였다.

경제 부문의 빅데이터 사업의 제목을 분석한 결과 여성문제와 관련이 있는 사업이 1개, 성별로 분리가 가능한 사업이 1개로서 전반적으로 여성과 관련된 빅데이터의 생산은 매우 저조한 상황임을 알 수 있다.

여성문제와 관련이 있는 사업은 ‘젠더 분리 관련 재산권(Gender disaggregated property ownership)’ 사업이다. 이 사업은 젠더적 관점에서 여성과 남성의 재산권에 관한 자료 생산을 목적으로 하는 빅데이터 사업이다(부록 1 참조). 이 사업은 지속가능한 발전 목표(sustainable development goal) 중 양성평등이라는 지표를 사업 소개에 담고 있다.

그 외에 경제부문의 빅데이터 사업 중에 성별 분리가 가능한 사업으로는 ‘센서스자료와 상업적 데이터의 비교(Comparing marketing databases with census data)’ 사업이 있다. 이 사업은 인구센서스 자료와 실제 상업 마케팅 데이터를 통하여 여성과 남성의 인구수를 비교하는 연구로서 빅데이터의 성별 분리의 중요성을 보여준다.

영국의 한 연구는 서베이 자료와 행정자료 간의 결합을 통한 빅데이터의 생산으로 성별 소득의 차이를 보여주고 있다.³²⁾ 개인 수준의 소득 관련한

30) <http://unstats.un.org/bigdata/>

31) <http://unstats.un.org/bigdata/inventory/>

32) Jack Britto, etc. 2015. Comparing sample survey measures of English earnings of graduates with administrative data during the Great Recession. The Institute for Fiscal Studies (IFS).

정부의 행정 자료를 활용한 빅데이터 분석 결과에서 대학을 졸업한 여성이 대학을 졸업하지 않은 여성들보다 중위소득이 약 3.5배 정도 높은 반면 남성인 경우 그러한 비율이 2.2배 정도인 것으로 나타났다. 기존의 추정치를 근거로 하는 서베이 데이터의 제한점을 해결할 수 있는 신뢰성 높은 근거 자료로 정책 활용에 사용될 수 있는 자료로 간주될 수 있다. 실제로 동 연구에서 영국의 기존 서베이 자료와 비교연구 결과를 살펴보면, 기존 서베이 자료를 활용한 젠더 차이 연구 결과 보다 빅데이터 자료 분석 결과의 젠더 차이가 작다는 연구결과를 도출했으며 시계열 분석에서도 지속적으로 같은 연구결과를 보고하고 있다.

나. 안전부문 활용사례

앞에서 언급된 UN 빅데이터 글로벌 워킹 그룹 사이트에서 비 경제 영역 부문의 속한 빅데이터 사업 내용을 분석하였다. 분석 결과 여성차별과 관련한 사업을 발견하였다.

소셜 미디어를 활용하여 직장에서 여성 차별 관련 패턴 인지(Feasibility Study: Identifying Trends in Discrimination Against Women in the Workplace in Social Media) 사업은 지속가능한 발전 목표(sustainable development goal) 중 양성평등(gender equality)이라는 목적을 포함하고 있었다.

동 사업의 내용은 빅데이터와 2030 지속적 발전을 위한 아젠다 관련 보고서³³⁾ 내용 중 소셜 미디어를 활용해 인도네시아 직장 내 여성에 대한 차별을 추적하였다. 동 사업은 인도네시아의 직장 내 여성근로자에 대한 차별이 심각함을 인지하고 국제노동기구(ILO)와 협력하여 빅데이터 관련 연구가 시작되었다. 2010-2013년 기간 동안 개인 정보 보호를 보장한 상태에서 기업의 사회 공익 추구라는 이해와 협조아래 직장에서 여성 차별 관련 빅데이터 활용 데이터를 구축하게 되었다. 데이터 구축에 따른 여성차별 관련

33) ABBAS MAAROOF, 2015. BIG DATA AND THE 2030 AGENDA FOR SUSTAINABLE DEVELOPMENT: FINAL DRAFT REPORT OPEN FOR COMMENT. UN; Global Pulse, 'Feasibility Study: Identifying trends in Discrimination against Women in the Workplace in Social Media', Global Pulse Project Series no 11, 2014'

분석이 이루어졌으며 이는 다시 보다 나은 여성가족 정책으로 환류 될 수 있다. 트위터 및 온라인 데이터를 활용하여 여성에 대한 직장 내에서의 성희롱 사건 및 직장 내 여성에 대한 차별을 실시간 모니터링 함을 목적으로 하고 있다. 기존의 직장 내 여성 차별 관련 설문 조사는 시간과 비용이 많이 드는 제한점이 존재하는 반면 온라인 데이터 실시간 분석은 비용 절감 및 실시간 여성에 대한 차별 현황 분석을 장점으로 지적하고 있다.

동 사업은 트위터 및 온라인 상에서 1) 남편으로부터 직장 취업 사전 허가(permission to work) 2) 직종과 일에 따른 여성 취업 적합성(appropriateness of work), 3) 일과 가정에서의 여성 근로자 부담(the burdens of working women), 4) 미모의 여자 근로자 선호 등 취업자격 관련 차별(Discrimination of job requirements), 5) 직장에서 여성 근로자 기술 및 교육의 부족(lack of skills or education), 6) 여성 근로자 취업하기 위한 비 공식적 비용 지불(cost of access employment), 7) 노동력 및 임금 착취 관련 여성 가사 노동자(home-based workers), 8) 직장에서 성희롱(sexual harassment in the workplace)을 포함하는 트위터 상의 내용을 8가지의 주제로 구분하여 직장에서 차별 패턴을 분석하였다.³⁴⁾

동 사업의 주요 결과³⁵⁾를 살펴보면, 첫 번째로 남편으로부터 직장 취업 사전 허가(permission to work)와 관련하여 3,000건의 트위터 정보를 수집하였다. 트위터 내용 중 “지난 번 직장은 남편이 그만 두어라고 해서 그만두었고 지금의 새 직장도 남편의 동의를 구했어”라는 온라인상의 트위터 내용을 사례로 제시하였다. 두 번째로, 직종과 일에 따른 여성 취업 적합성(appropriateness of work)과 관련하여 5,000건의 트위터 정보를 수집하였다. 트위터 내용 중 “나에게 이 일을 시키지마! 여성근로자들이 밤에 일하면 자궁경부암에 걸릴 가능성이 높아”라는 온라인상의 트위터 내용을 사례로 제시하였다. 세 번째로, 일과 가정에서의 여성 근로자 부담(the burdens of working women)과 관련하여 21,000건의 트위터 정보를 수집하였다. 트위터

34) UN Global Pulse, 'Feasibility Study: Identifying Trends in Discrimination Against Women in the Workplace In Social Media', Global Pulse Project Series no.11, 2014'. 2쪽.

35) UN Global Pulse, 'Feasibility Study: Identifying Trends in Discrimination Against Women in the Workplace In Social Media', Global Pulse Project Series no.11, 2014'. 2쪽.

IV. 해외의 여성가족 빅데이터 활용전략 및 사례 ●●● 59

내용 중 직장과 가정에서의 일로 인한 여성 근로자의 부담에 대한 걱정으로 “일을 할 거야 아니면 가정주부로 돌아갈 거야”라는 온라인상의 트위터 내용을 사례로 제시하였다. 네 번째로, 미모의 여자 근로자 선호 등 관련 취업 자격 조건 관련 차별(Discrimination of job requirements) 관련 78,000건의 트위터 정보를 수집하였다. 관련 트위터 내용 중 “접대원(receptionist) 구인 광고 자격 조건; 1. 여성일 것, 2. 매력적일 것, 3. 최소 고등학교 졸업, 4. 23세 이하”라는 온라인상의 트위터 내용을 사례로 제시하였다. 본 사업 후 이러한 4가지 주제에 대해 심층적인 연구가 필요함을 제기하였고 직장에서의 실시간 모니터링 및 분석 결과에 근거하여 직장에서의 여성 차별 관련 빅데이터 활용 가능성을 높인 것으로 나타났다.

다. 보건 부문 활용 사례

앞에서 언급된 UN 빅데이터 글로벌 워킹 그룹의 사업 중 건강증진 사례 이면서 성인지적인 사업은 위생청결 관련 대중 인지에 대한 소셜미디어 분석 (Analyzing Social Media Conversations to Understand Public Perceptions of Sanitation)으로 나타났다. 이 사업은 위생청결 관련 트위터 내용 분석을 실시하였다. 트위터 내용 중 콜레라와 관련된 대중의 인지 속에 젠더 이슈가 위생청결과 연관되어 나타난 것이 이 사업의 중요한 결과로 나타났다. 또한 성인지적인 내용 사업으로 소셜데이터를 활용하여 피임과 청소년 임신에 대한 태도에 대한 분석(Analyzing Attitudes Towards Contraception & Teenage Pregnancy Using Social Data)이 존재하였다(부록 2 참조).

일리노이 주의 Better Birth Outcomes (BBO) 프로그램³⁶⁾은 고위험 임신부 대상 보다 효과적이고 집중적인 지원을 받을 수 있게 돕기 위해 계획되었고 빅데이터가 활용되었다. 미국의 임신부 중 고위험 임신 및 출산 예방을 위하여 빅데이터의 활용은 공공자료의 연계 측면에서 주목할 만하다. 일리노이주에 거주하는 프로젝트 참가자 73만 케이스의 출산에 대한 자료를 수집

36) Laura Nolan & Rashida Brown. IDHS Project: Better Birth Outcomes
<https://dssg.uchicago.edu/2015/07/17/idhs-project-better-birth-outcomes/>

60 ●●● 여성가족 빅데이터 활용을 위한 기초연구

하고 활용하였다. 위험한 출산 결과가 나타날 상황을 더 잘 이해하기 위해 개인별 임신 관련 정보와 인구조사 자료와 연계하여 분석하였다. 빅데이터로 형성된 자료들의 내용은 광범위하며 산모의 이전 임신, 사례별 사회복지사와의 관계, 임신 기간 동안의 건강관리 및 출산 결과에 대한 정보를 포함하고 있다. 스트레스, 사회경제적 지위, 술, 담배 또는 약물 사용, 영양 부족, 불충분한 건강관리 그리고 산모의 나이 등의 고위험 출산 요인을 빅데이터로 분석하였다. 분석 방법으로 다양한 통계수치, 장비 기계 학습 및 매핑 기술을 이용하고 산학연 협동으로 고위험 임신 및 출산에 처한 여성을 위한 프로그램을 개선하였다. 또한 지리적인 요인에 분석의 초점을 맞추고 BBO 프로그램 등록 과정 및 가장 위험에 처한 산모들에게 자원을 효과적으로 할당하는 프로그램은 산모와 태아의 건강을 크게 개선할 수 있을 것이다.

V

결론 및 정책제언

1. 결론	63
2. 정책제언	67



1. 결론

현대 사회에서 데이터의 중요성은 날로 높아져가고 있다. 기업은 물론 정책 결정자들과 연구자들도 데이터에 기반 한 결정이 필수적이 되고 있다. 빅데이터는 이러한 데이터의 활용과 분석에서 핵심적인 개념으로 자리잡고 있다. 여성가족 정책 수립과 연구 역시 지금까지 성인지 통계의 생산과 활용에만 의존하였던 것에서 탈피하여 다양한 빅데이터를 이용한 연구결과를 도출함으로써 기존에 하지 못했던 사각지대를 커버할 수 있을 것이다.

본 연구는 여성과 가족 분야에 대한 연구에서 빅데이터가 활성화되지 못한 현 시점에 빅데이터를 어떻게 활용할 수 있을 것인가에 초점을 맞추고 진행하였다.

가. 여성가족 빅데이터의 정의와 범주

빅데이터의 처리과정의 특성상 데이터가 초기부터 통계분석에 적합한 형태로 있기 보다는 복잡하고 다양한 분석과정을 통해 빅데이터로서의 역할을 하게 된다. 따라서 여성가족 빅데이터라는 개념은 사후적으로 여성과 가족에 대한 연구에서 사용된 빅데이터를 지칭하는 것이 적합하다. 그러나 이러한 귀납법적인 접근방식으로는 빅데이터를 활성화 하는데는 한계가 있다. 본 연구에서는 이러한 문제점을 고려하여 여성가족 빅데이터를 단정적으로 정의하기 보다는 어떤 성격의 자료들이 여성가족 빅데이터로 활용가능성이 높은지를 탐색하였다.

첫째 여성가족 빅데이터로 활용가능성이 높은 자료들은 여성정책의 수행과 연구를 위해 지금까지 발생된 자료들이다. 이에 해당 것으로는 여성가족부와 한국여성정책연구원, 지방자치단체 여성가족 관련 부서, 지방의 여성정책 연구소 등에서 축적된 자료들이 있다. 국가승인통계, 행정업무에서 발생한 통계, 각종 행정자료, 시설 및 센터에서의 실적과 상담자료, 복지사업에서의 대상자관련 자료, 홈페이지와 해당 페이지 로그인 자료 등을 주제에 맞추어 결합할 경우 빅데이터를 구축하여 새로운 연구결과를 얻을 수 있다.

두 번째 범주는 빅데이터의 생산과 활용에 성인지적 관점을 포함하거나,

성별 분리를 시도함으로써 발생하는 데이터를 예상할 수 있다. 예를 들어 서울시의 “서울시 정책지도”나 “우리마을가게 상권분석서비스”는 기초적인 행정자료와 지리정보 등을 이용하여 빅데이터를 분석한 획기적인 정책서비스이다. 이와 같이 빅데이터의 구축에 있어서 성별차이를 보여주거나, 여성이슈 등을 고려한 내용을 포함한다면 여성가족 빅데이터의 범주로 활용할 수 있을 것이다.

마지막으로서는 기존의 빅데이터 자료 유형들에 대한 성별로 분리하거나 여성 주제를 반영할 경우 여성가족 빅데이터의 활용이 가능할 수 있다. 예를 들어 신용/직불카드 금융자료 등이 현재 주요 빅데이터 자료로 활용되고 있다. 성별 이슈를 반영하여, 신용불량자가 되는 원인이나 유형, 패턴 등을 신용카드 거래자료와 개인별 소득자료, 가구소득 자료 등을 연결하여 본다면 남성과 여성의 신용불량자 유입 경로의 유사점과 차이점등을 분석할 수 있을 것이고, 성별로 적합한 예방책을 만들 수 있다.

나. 정부의 빅데이터 정책의 여성가족 관련성

정부의 빅데이터 정책을 보면 우선 범정부 정책으로 현 정부는 2011년 11월 교육과학기술부, 행정안전부, 지식경제부, 방송통신위원회, 국가과학기술위원회로 구성된 범정부 빅데이터 추진협의체를 구성하고 데이터의 창조적 활용을 통한 스마트 강국 실현을 비전으로 하는 추진전략과 목표를 수립했다. 추진과제로 6개 분야 빅데이터의 대상과제 16개를 제시하고 국민수혜, 유용성 등을 고려해 3개 우선추진과제로 선정하였다. 우선추진과제를 살펴보면 첫째 범죄발생 장소, 시간 예측을 통한 범죄발생 최소화, 둘째 예측 기반의 자연재해 조기 감지 대응 사업, 셋째 주민참여형 교통하고 감소체계 구축사업이다.

2013년 12월 관계부처 합동으로 수립한 빅데이터 산업 발전전략은 3개 분야의 8개 추진과제를 수립하였다. 이 전략은 국가차원의 빅데이터 ‘활용’을 촉진하고 빅데이터 ‘산업’을 육성하기 위한 전략적 지원이 필요한 시점임을 강조하고 있다.

주요 부처의 빅데이터 활용계획으로는 통계청, 국민안전처 등이 있다.

통계청은 2015년 10월 빅데이터를 활용한 통계작성을 위해 통계데이터허브국에 빅데이터통계과를 신설하였다. 빅데이터를 활용한 통계작성을 위한 제도적 기반을 구축하고 데이터 간 융·복합 사례를 발굴해 빅데이터 활용 확산에 나서고 있다. 특히 빅데이터의 연계와 융합에 가장 큰 걸림돌인 개인정보를 보호하면서 데이터 간 연계를 통해 빅데이터의 잠재가치와 활용에 공감대를 만들어간다는 기본방향을 갖고 있다. 특히 통계청은 데이터를 연계하고 분석 결과를 확산하는 역할을 하는 데이터 허브 기능을 수행할 계획이다.

국민안전처는 2016년 4월 전통적인 재난·안전 업무방식에 빅데이터를 접목 시키기 위해 재난예방 및 대응, 정책수립 지원, 활용기반 구축 등을 목표로 빅데이터 활용 종합계획을 수립했다. 이 계획에는 2020년까지 5년 동안 12개의 신규과제와 5개의 기존과제를 포함 총 17개 과제가 있다.

정부 주도의 공공 빅데이터 활용 사례로는 2014년 미래창조부 주관 하에 부처의 과제 공모를 통해 최종 4개 부처에 대해 컨설팅을 진행했다. 4개 부처의 하나인 여성가족부는 ‘13-‘14년도 새일센터의 구인기업, 구직자, 취업자 특성과 일자리 미스매치 원인에 대한 분석을 시도하였다.

국민건강보험공단은 국민건강·질병표준DB 구축사업을 수행하고 있다. 4대 보험의 통합 징수, 요양급여비용 지급 등 업무수행과정에서 축적된 자료를 재구축하여 빅데이터를 구축하였다. 부산광역시는 빅데이터를 활용하여 부산 도시서비스 분석 정보시스템을 구축하였으며, 인천광역시는 이동통신사의 빅데이터를 기반으로 내외국인 유입인구의 분석을 시도하였다. 통계청은 공공·민간 빅데이터 연계 현황 및 활성화 방안의 첫 번째 시도로 신혼부부에 대한 빅데이터를 구축하여 분석을 수행하였다.

공공부문에서 활용한 빅데이터 자료를 구분하면 크게 공공자료의 활용, 민간자료의 활용, 공공자료와 민간자료의 연계로 구분할 수 있다. 공공부문 빅데이터의 당면과제는 공공데이터와 개방과 개인정보 보호의 충돌이다. 개인정보 보호의 수집, 처리, 보유 절차가 엄격해짐에 따라 최소한 필요한 개인정보를 제외하고 폐기할 수 밖에 없는데, 이러한 개인정보의 폐기에 따라 개인의 성별 식별 정도가 같이 삭제되는 경우 성별 분석이 어려워진다. 개인정보 비식별화 과정의 필요성이 대두되고 있으며, 이 과정을 통해 개인의

식별은 불가능하지만 성, 연령 등의 속성은 유지되는 방향으로 빅데이터의 구축을 시도할 필요가 있다.

여성가족정책 빅데이터 연구를 위해서는 범정부 차원과 각 부처별 빅데이터 관련 세부과제를 통해 진행되는 사업의 경우 사업추진을 위해 활용하는 데이터의 성별 구분 정보와 이를 통해 생산되는 데이터의 성별 구분 식별 정보에 대한 검토가 필요하다.

다. 해외의 여성가족 빅데이터 활용 전략 및 사례

국제기구와 선진 국가들은 빅데이터를 이용한 정책개발과제 전략을 제시하여 강력하게 추진하고 있다. 미국의 “빅데이터 전략 수립 및 국가차원의 거버넌스 구축”이나 영국의 “정보공유·활용을 위한 데이터 개방중심의 정책” 등이 대표적인 사례이다. 그러나 각국의 빅데이터 정책에서 별도의 여성과 가족을 염두에 둔 빅데이터 전략이나 프로젝트는 찾기 어렵다.

이러한 상황에서 Data2X는 여성가족 빅데이터의 활용에 좋은 사례이다.

Data2X는 빅데이터 분석을 기반으로 전세계 여성과 소녀들에게 보다 나은 삶을 제공하기 위하여 2012년에 형성된 협력적 플랫폼으로 UN의 우선계획(Initiative)로 수행되고 있으며 William & Flora Hewlett 재단, Bill & Melinda Gates 재단 및 Clinton 재단이 지원하고 있다. Data2X이 빅데이터 분석의 기준은 UN이 제시한 지속가능한 발전 목표(Sustainable Development Goals, SDGs)에 대한 성별영향평가이다. 이를 위하여 UN 산하기관, 정부, 시민단체, 학계 등과 협력하여 젠더 데이터를 구축, 확장, 분석하여 효과적인 정책과 전략을 제공하고 의사결정을 제시한다. Data2X는 빅데이터 분석을 위하여 필요한 자료를 확인하면서 성별 데이터 차이(gender data gap) 변수를 확인하여 이를 건강, 교육, 경제기회, 정치참여, 안전의 다섯 가지 영역에서 28가지 변수로 분류하고 있다.

이 외에도 Data2X는 빅데이터 분석을 통하여 전세계 여성건강을 도모하기 위한 소규모 프로젝트를 진행하고 있다(Data2X, 2015). 빅데이터를 이용한 성별분석 소규모 프로젝트는 휴대폰 이용 기록, 위성 데이터, 소셜미디어를 이용한 프로젝트가 있다.

Data2X가 갖는 중요성은 첫째, 여성의 성평등을 위해 데이터의 격차를 줄여야하고 이를 위해 빅데이터의 활용이 필요하다는 인식을 갖고 있는 프로그램이라는 점이다. 둘째, 직접 빅데이터를 생산하기 보다는 다른 기관과 협력체계를 구축하여 여성 빅데이터의 생산을 촉진하고 있다는 점이다. 즉 여성이나 가족을 별도로 분리하여 자체적으로 빅데이터의 구축을 시도하기 보다는 다른 기관과의 연계를 통하여 효율적으로 생산하고 있는 점이 큰 시사점이다.

UN의 빅데이터 글로벌 워킹 그룹은 상당히 많은 양의 데이터 사업을 소개하고 있다. 이 사업 중 여성이나 가족과 관련된 내용으로는 “성별 분리된 재산권실태”, “소셜 미디어를 활용한 직장에서 여성 차별 관련 패턴의 인지”, “위생청결 관련 대중 인지에 대한 소셜 미디어 분석”, 소셜 데이터를 활용한 “피임과 청소년 임신에 대한 태도 분석” 등이 있었다.

일리노이 주의 Better Birth Outcomes 프로그램은 고위험 임산부들이 보다 효과적으로 집중적이고 집중적인 지원을 받을 수 있도록 돕기 위해 빅데이터를 활용하였다.

2. 정책제언

지금까지의 연구결과에 의하면 여성가족 빅데이터에 대한 정부의 정책이나 계획이 전무한 상태로 여성가족 연구에서 빅데이터의 활용은 여러 어려움과 불편함이 있다. 빅데이터 연구의 활성화를 위해 다음과 같은 정책을 제언한다.

가. 활용계획의 수립

여성가족 빅데이터의 활용에 대한 국가 전략과 데이터 거버넌스를 포함한 활용계획의 수립이 필요하다. 이 계획을 위해서는 사전에 여성가족 데이터의 빅데이터 활용 가능성과 빅데이터 활용 사업의 개발, 기술 인프라의 구축 등에 대한 사전 연구가 필요할 것이다. 이 계획에는 법, 제도와 데이터의 연계 공유에 대한 가이드라인, 활용을 위한 제도 개선 등이 포함되어야 할 것이다.

나. 여성가족정책 빅데이터의 원천자료 선별과 표준화

여성가족정책의 실행과 연구에서 빅데이터의 이용을 활성화하기 위해서는 사전에 빅데이터의 가능성이 높은 자료를 선별하여 이들 자료의 축적이 선행되어야 한다. 여성가족부와 산하 기관 및 연구기관을 중심으로 여성가족정책 관련 조사통계와 행정통계 및 사업실행 과정에서 시스템적으로 발생하는 정보를 총체적으로 점검하여 활용가능성과 필요성이 높은 자료들을 빅데이터로 활용할 수 있도록 사전 작업이 필요하다. 선별된 자료들은 데이터 축적과 연계를 고려한 표준화 작업을 거쳐 빅데이터로의 활용성을 높인다. 특히 현재 빅데이터의 활용의 걸림돌이 되고 있는 개인정보 보호를 준수하면서 자료간의 연계가 가능하도록 하는 비식별화 과정에 대한 철저한 준비가 필요하다.

다. 여성가족정책 빅데이터의 DB 구축 및 운영

여성가족에 대한 빅데이터 연구가 진행되고 데이터가 생산되기 시작하면 관련 정보를 제공하는 시스템의 필요성이 생긴다. 빅데이터의 성격상 모든 자료를 다 제공하기는 불가능하지만, 빅데이터의 최종 분석결과 이전 단계의 자료에 대한 수요가 커질 것이다. 즉, 해당 빅데이터에서 활용하고자 하는 키워드를 선정하면 분석 가능한 형태의 데이터로 제공하는 시스템의 요구가 커질 것이며 이에 대한 준비가 필요하다.

예를 들어, 한국언론진흥재단에서 서비스하고 있는 뉴스 빅데이터 분석 시스템인 BigKinds(빅카인즈)를 들 수 있다(<http://bigkinds.or.kr>). 이 시스템은 주요 언론사의 뉴스를 수집하는 시스템을 통해 형태소 분석, 개체명 분석, 네트워크 분석 등 기사를 분석하는 뉴스분석 시스템을 구축하고 있다. 뉴스를 검색할 경우 뉴스 히스토리, 관계도, 언급추이 등 시각화한 다양한 차트를 함께 제공하고 있다. 빅 카인즈 프로(Big Kinds-Pro) 서비스는 더 전문화한 것으로 특정 뉴스에서 정보를 추출할 수 있도록 개인화 된 텍사노미 관리, 질의어 관리, 정보 추출 패턴 규칙 적용 기능을 제공하고 트렌드, 네트워크 분석 등 다양한 분석도구를 통해 그 결과를 활용할 수 있다.

라. 정부 빅데이터 사업에 대한 성인지적 개선과 활용

국가승인통계의 신규 승인 시에는 통계법에 의해 특별한 사유가 없는 한 모든 개인차원의 통계는 성별 분리하여 조사, 발표하도록 되어 있다. 그러나 빅데이터는 공공부문에서 생산하더라도 성별 분리에 대한 의무조항이 없어, 성맹적(sex-blinded)으로 생산되고 있거나, 생산될 가능성이 높다. 범정부 차원 또는 각 부처별 빅데이터 관련 세부과제를 통해 진행되는 사업에서 성별 이슈가 있더라도 성별 분리가 되지 않게 수집, 가공 된다면 여성가족 정책을 위한 빅데이터를 활용하기가 어려워진다. 특히 개인정보 보호를 위해 개인 정보의 수집 시 주민등록번호의 뒷부분 7자리를 삭제하는 경우가 있는데, 이때 성별 정보가 소실되므로 이러한 일이 발생하지 않도록 신속한 대책이 필요하다.

빅데이터를 활용한 데이터 생산과 분석의 경우 기존에 축적되거나 생산되고 있는 데이터를 활용해 분석하거나 연계해 활용하고 있기 때문에 활용하는 정보에 개인의 성별구분 정보가 포함되어 있는 경우 굳이 분석항목에서 제외해 분석할 필요가 없어진다. 그러나 빅데이터의 구축 시에도 개인정보가 있는 통계와 다른 정보의 연계 시에는 개인비식별화 작업을 거치게 되는데 마찬가지로 성별 정보가 탈락될 가능성이 있으므로 이에 대한 대책이 필요하다. 또한 사용한 데이터가 성별구분이 불가능한 경우 왜 불가능한지에 대한 정보를 제공한다면 향후 관련 데이터 수집에서 성별구분 정보를 수집할 수 있는 방법과 데이터 수집 개선 방안의 기초자료로 제공될 수 있다는 점에서 큰 의의를 갖는다.

마. 여성가족 빅데이터 연구에 대한 정책지원

여성가족 빅데이터를 활성화하기 위한 기초연구에 대해 정책적 지원이 이루어져야 한다. 필요한 연구주제로 첫째 여성가족정책 관련 빅데이터 생산 및 연계방안 연구가 필요하다. 빅데이터 연구가 필요한 여성가족정책 연구의 주제를 발굴하고 빅데이터 생산방안을 검토하는 연구의 필요성이 제기된다. 생산의 개념은 새로운 데이터의 발굴 뿐 아니라 기존 데이터 연계를 포함

하는 포괄적인 의미로 개별 부처 내에서 생산하는 데이터의 연계에서부터 부처 간 데이터의 연계, 공공데이터와 민간데이터의 연계까지 범위를 확장해 생산방안에 대한 종합적인 검토가 진행되어야 할 것이다. 두 번째 주제는 여성가족정책 관련 국가승인통계 중 빅데이터와 연계하거나 대체가 가능한 조사통계에 대한 검토 및 발전방안을 수립하는 연구이다. 정부 정책의 의사결정에서 데이터의 중요성을 점차 증대하고 있고 데이터 기반 환경에서 통계의 중요성은 더욱 강조 되고 있으나 조사통계의 경우 조사거부와 불응 등 응답률이 저하되는 문제로 인해 조사 환경이 악화되고 있다. 1인 가구의 증가와 가구원 수의 감소로 주간 방문 시 응답률이 저하되고 있고 프라이버시 보호가 강화되는 사회적 분위기 속에서 최초 대면을 위한 접근이 차단되는 사례가 매년 증가하고 있다. 이러한 사회적 분위기 속에서 여성가족정책과 관련된 조사통계의 어려움은 더욱 높아질 것이 자명해 보인다. 특히, 경력단절, 다문화, 폭력, 성매매 등 민감성이 높은 조사가 상대적으로 많은 여성가족부의 상황을 고려할 때 데이터 연계를 통해 조사를 대체하거나 응답 부담을 줄여 응답률을 높여나갈 수 있는 방안에 대한 검토가 요구된다고 할 수 있다.

참고문헌

- 국민건강보험(2015). 지역단위 맞춤형 건강관리를 위한 의료이용지표 활용 매뉴얼
- 국민안전처(2016). “빅데이터 활용 종합계획 수립” 보도자료
- 관계부처 합동(2013). 창조경제 및 정부3.0 지원을 위한 빅데이터 산업 발전전략
- 교육과학기술부, 행정안전부, 지식경제부, 방송통신위원회, 국가과학기술위원회 (2012). 스마트 국가 구현을 위한 빅데이터 마스터플랜
- 이지영(2015). “빅데이터의 국가통계 활용을 위한 기초연구”. 통계개발원. 『2015년 상반기 연구보고서 제Ⅱ권』
- 정용찬(2012). “빅데이터 혁명과 미디어 정책 이슈” KISDI Premium Report 12-02.
- 정용찬(2013). 『빅데이터』 커뮤니케이션북스.
- 채승병·안신현·전상인(2012). “빅데이터: 산업지각변동의 진원” CEO Information 2012.5.2. 제 851호.
- 통계청(2016). 공공·민간 빅데이터 연계 현황 및 활성화 방안 - 신혼부부 통계를 중심으로 보도자료
- 한국정보화진흥원(2014). 2014년 공공 빅데이터 분석사업 최종 결과물
- 한국정보화진흥원(2014). 데이터 기반의 미래전략 컨설팅 완료보고서 - 여성가족부
- Abbas Maarooof, 2015. Big Data And The 2030 Agenda For Sustainable Development: Final Draft Report‐Open For Comment. 9쪽.
- Abbas Maarooof, 2015. Big Data And The 2030 Agenda For Sustainable Development; Global Pulse, ‘Feasibility Study: Identifying trends in Discrimination against Women in the Workplace in Social Media’, Global Pulse Project Series no 11, 2014’
- Buvinic, M., Furst-Nichols, R., & Koolwal, G. (2014). Mapping gender data gaps, UN Foundation. Retrieved August 30 from http://data2x.org/wp-content/uploads/2014/11/Data2X_MappingGenderDataGaps_FullReport.pdf
- Data2X. (2014). Ready to Measure Progress on Gender Equality. Retrieved from <http://data2x.org/what-is-gender-data/ready-to-measure>
- Data2X. (2015). Big Data & Gender Data Gaps. Retrieved from <http://data2x.org/wp-content/uploads/2014/08/Big-Data-Projects.pdf>
- Global Partnership for Sustainable Development Data (2016). Data collaboratives. Retrieved from <http://www.data4sdgs.org/dc-data2x>

- Laney Douglas(2001). “3D Data Management Controlling Data Volume, Velocity, and Variety” Meta Group, Application Delivery Strategies,
- Jack Britto, et. al. 2015. Comparing sample survey measures of English earnings of graduates with administrative data during the Great Recession. The Institute for Fiscal Studies (IFS).
- Laura Nolan & Rashida Brown. IDHS Project: Better Birth Outcomes
<https://dssg.uchicago.edu/2015/07/17/idhs-project-better-birth-outcomes>
- LeFurgy, William G. (2015). Stewarding big data: Perspectives on public access to federally funded scientific data: In the “Big data, big challenge in evidence-based policymaking edited by Jaysuriya, H. Kurmar and Ritcheske, Kathryn A. 3-16쪽.
- Rajiv Niles(2015), “Data & Analysis in the Internet of Everything” 매일경제신문, 미래창조과학부. 「2015년 빅데이터컨벤션 및 성과보고회」
- The Federal Big Data Research And Development Strategic Plan. 2016. The Networking And Information Technology Research And Development Program.
- UN Global Pulse, 'Feasibility Study: Identifying Trends in Discrimination Against Women in the Workplace In Social Media', Global Pulse Project Series no.11, 2014'. 2쪽.

관련사이트

- 서울시 빅데이터 활용 정책지도 홈페이지. <http://gis.seoul.go.kr>
- 여성가족부 정부 3.0 페이지.
http://www.mogef.go.kr/korea/view/policy/policy04_01_01a.jsp
- 여성가족부 공공데이터 개방 현황.
http://www.mogef.go.kr/korea/view/policy/policy04_11.jsp
- 우리마을가게 상권분석서비스 홈페이지. <http://golmok.seoul.go.kr>
- 한국언론진흥재단 뉴스 빅데이터 분석 시스템 빅카인즈. <http://bigkinds.or.kr>
- Big Data Project Inventory. <http://unstats.un.org/bigdata/inventory>
- Global Working Group(GWG) on Big Data. <http://unstats.un.org/bigdata>



부 록

- 부록 1. 경제 영역 부문 빅데이터 사업:
UN 빅데이터 글로벌 워킹 그룹
(Big data global working group) 75
- 부록 2. 비 경제 영역 부문 빅데이터 사업:
UN 빅데이터 글로벌 워킹 그룹
(Big data global working group) 78



부록 1. 경제 영역 부문 빅데이터 사업: UN 빅데이터 글로벌 워킹 그룹 (Big data global working group)

사업명	젠더 관점 여부	성 분리 여부
Feasibility study on web scraping for labor market indicators	X	X
Feasibility study on the use of smart meters: using electricity, gas and water smart meter data for energy and environmental statistics	X	X
CPI with Scanner Data	X	X
Automatic price collection on the internet: Using web scraping and web Crawlers for price index compilations	X	X
A Big Data Pilot Project – With Smart Meter Data	X	X
Using web scraping price data for price index of e-commerce	X	X
The comparison between the data of interbank transactions and the retail sales: credit card data for use in verifying retail sales	X	X
Using scanner data for national accounts, household income and expenditure and business statistics	X	X
Using scanner data for compilation of CPI	X	X
Use of mobile phone data for mobility and urban statistics	X	X
Using Google Trends for "now-casting" unemployment	X	X
Flight reservation systems	X	X
Multipurpose consumer price statistics: sub-project Scanner Data	X	X
Web scraping for price statistics	X	X
Web scraping data from retailers' websites for the CPI calculation	X	X
Estimate the job vacancies based on information from job portals, using web scraping tools.	X	X
Online cash register data from retail stores for retail trade statistics	X	X
Scanner data for compiling the CPI	X	X
Use of Scanner Data for CPI	X	X
Internet as a data source on ICT in enterprises	X	X
Use of Google Trend for "now-casting" the unemployment rate	X	X
Web scraping and scanner data for price statistics	X	X

76 ●●● 여성가족 빅데이터 활용을 위한 기초연구

사업명	젠더 관점 여부	성 분리 여부
Use scanner data for price statistics	X	X
Price information based on scanner data and website information	X	X
Exploring the use of social media messages for economic indicators	X	X
The use of online data in the HICP	X	X
Online Price Changes of Means of Production	X	X
Big Data Enterprise Statistical Indicator	X	X
Using web scraping data for job offers statistics	X	X
Big Data Sandbox web scraping project	X	X
Online price index	X	X
Using scanner data for price and economic statistics	X	X
Testing web scraping tools and scanner data for price statistics	X	X
Assessing use of scanner data for compiling the Consumer Price Index	X	X
Web scraping of prices for the CPI	X	X
Possible improvements of Household Budget Survey using scanner and credit card data	X	X
Use of scanner data for price statistics	X	X
Price collection with scanner data	X	X
Inclusive infrastructure for sustainable development	X	X
Comparing marketing databases with census data	X	○
Comparing marketing databases with census data		
Price collection with scanner data	X	X
U.S. household food purchases and retail food sales	X	X
Evaluation of credit card data for use in estimates of consumer spending	X	X
Non-probability online web panels	X	X
Climate impacts on agriculture: using weather and climate data and satellite imagery for rural and resource economics	X	X
Web Scrape and Application Program Interface (API): exploring webscraping on price data	X	X
Autocoding: using payroll systems data for labor market statistics	X	X

사업명	젠더 관점 여부	성 분리 여부
Real-time price discovery in commodity futures markets	X	X
Innovation Measurement Initiative	X	X
Pilot Study to Capture Granular High-Frequency Price Data Using Modern Information and Communication Technologies	X	X
Using mobile phone data for national, sub-national and geo-coded average prices	X	X
Gender disaggregated property ownership	○	○
How good are CDR-derived measures of income and inequality, and can governments systematically use them?	X	X
Targeting Poverty by Predicting Poverty: Using Machine Learning in Targeted Transfer Programs	X	X
Big Data for Financial Inclusion and Poverty Mapping	X	X
Forecasting Poverty and Shared Prosperity Using Cell Phone Data	X	X
Real Time Assessments of How Markets Are Working for the Poor	X	X
Real Time Forecasting of Skills Demand and Supply: Analytics of Big Data from Babajob in India	X	X
Big Data solutions for enhancing tax compliance	X	X
"Now-casting" Food Prices in Indonesia Using Social Media Signals	X	X
Big Data tool to conduct automated roof counting to monitor poverty	X	X
Modernization of price collection and compilation with web and scanner data	X	X

부록 2. 비 경제 영역 부문 빅데이터 사업: UN 빅데이터 글로벌 워킹 그룹(Big data global working group)

사업명	젠더 관점 여부	성 분리 여부
Tourism statistics produced from camera data to estimate the incoming and/or outgoing traffic at the borders of Hungary	X	X
Improving social transparency in road safety and maintenance operations in Albania through ICT interfaces	X	X
Big Data "Just in time analytics" in disaster risk management activities	X	X
Global Urban Monitoring Using High Resolution Night-time Lights	X	X
Using data from cell phone networks to measure rainfall in data scarce context	X	X
Telecommunications project: using mobile phone roaming data to produce tourism statistics	X	X
Smart Metering Data: using smart meter data on electricity usage for housing and demographic statistics	X	X
UNECE Wiki Project: studying correlations between wikistats and official statistics	X	X
Transport Statistics from mobile phone data and camera data	X	X
Persons and Places: Mobility Estimates based on Mobile Phone Data	X	X
Use of data from social networks to obtain statistical and geographical information	X	X
Use of satellite imagery to obtain geographical information	X	X
Traffic intensities statistics: use of road sensors	X	X
Daytime population statistics based on mobile phone location data	X	X
Feasibility study on geo-localization: using geographical data from web services for geocoding static objects	X	X
Capacity Building in using Big Data as sources for public statistics	X	X
Capacity building for the use of Big Data for statistical purposes	X	X
Non-Residential Buildings Inventory: Feasibility Study	X	X
A set of indices of accessibility and remoteness	X	X
Crop survey by farmland: using satellite and aerial remote sensing to help estimate agricultural statistics	X	X
The application of Big Data for highway and waterway transport statistics	X	X

사업명	젠더 관점 여부	성 분리 여부
Pilot project on the use of mobile positioning data for tourism statistics	X	X
Pilot project using social media to create a happiness index	X	X
Feasibility study on creating indicators using web scraping	X	X
Developing a Curriculum and Training Modules on Using Big Data for Official Statistics	X	X
Wikistats: Exploring Wikipedia page view counts as a source for official statistics	X	X
Analysis of methodologies for using the Internet for the collection of information society and other statistics	X	X
Feasibility study on the use of mobile positioning data for tourism statistics	X	X
Using traffic sensor data targeting for commuting times and distance	X	X
Feasibility study on the use of mobile telephone data for tourism & transportation statistics	X	X
Training Institutions Frame Creation: using web scraping to improve survey frames	X	X
Migration: using mobile phone data for migration statistics	X	X
Using Big Data sources in official statistics	X	X
Daily migration of population: using mobile CDR data for daily migration data	X	X
Integrated Environment System (IES): using environmental sensing systems and data analytics for real-time environmental information	X	X
Population Estimates: using administrative data from many sources for population estimates	X	X
Merging statistics and geospatial information	X	X
Big Data for the estimation of origin-destination mobility matrices	X	X
Evaluation of automatic identification system (AIS) data as a source for improved and new statistics	X	X
Identifying sources and uses of webscraping	X	X
Monitoring SDG 16 on peace and justice through Big Data	X	X
Log Analysis: becoming more familiar with Big Data technologies	X	X
Smart meter type data for household structure/size and occupancy	X	X
Smart meter data potential for detecting unoccupied dwellings	X	X

80 ●●● 여성가족 빅데이터 활용을 위한 기초연구

사업명	젠더 관점 여부	성 분리 여부
Secondary Source, purchased data: exploring the purchase of data from secondary sources	X	X
Use of electronic health records	X	X
Administrative Data Sources	X	X
Health Care Satellite Account: using health records for health spending and health care prices	X	X
Physical activity data: using physical activity monitors to collect physical activity data	X	X
Patent Technology Classifications	X	X
Computerized Criminal History Records	X	X
Genetic data	X	X
Arrest-Related Deaths: using Google Alerts and direct survey methods to measure arrest-related deaths	X	X
Using administrative data to augment compensation surveys	X	X
A Pilot Study To Establish Data Use and Quality Standards for Using New Sources of Structured and Unstructured Data	X	X
Central America Big Data project	X	X
Big Data for Freight Transport and Logistics Policy Making	X	X
How Good Are CDR-Derived Measures of Income and Inequality, and Can Governments Systematically Use Them?	X	X
Using Big Data Analytics to Discover Patterns of Medical Insurance Utilization for Medical Cost Monitoring in China	X	X
Analyzing data to identify indicators of corruption, fraud, or collusion in Bank-financed projects or public spending	X	X
Predicting vulnerability to flooding and enhancing resilience using big data	X	X
Data diving and sprinting – making the case for predictive analytics at the Bank	X	X
Evaluating crime and infrastructure in Bogotá, Colombia	X	X
Geospatial country aid data comparisons to global aid disbursements	X	X
Open Transport Data and Urban Mobility Lab in Sao Paulo – Improving urban mobility using technology and data analysis	X	X
Improving Procurement Outcomes through Data Analytics	X	X

사업명	젠더 관점 여부	성 분리 여부
Innovations in Big Data Analytics for Development	X	X
Knowledge Management Product on "Data Analytics for Intelligent Energy"	X	X
Investment Lending Operation - Jamaica Energy Security and Efficiency Enhancement Project	X	X
Supporting the IFC EDGE team to demonstrate examples of building energy efficiency interventions in Green Certified Buildings	X	X
Smart Agriculture in Central America: A Big Data Roadshow Central America	X	X
Surfacing and Incentivizing Opportunities to Advance Big Data Analytics	X	X
Using satellite imagery to get a clearer sense of population and assets location	X	X
Spatial allocation of population (WorldPop) and built-up areas can be used to calculate densities, etc.	X	X
Classify NTM Using Machine Learning	X	X
Using Big Data to Measure City Innovation Capacity	X	X
Leveraging Big Data to Help Competition Agencies Tackle Anti-Competitive Behavior	X	X
Big Data for Climate Smart Agriculture - Enhancing & Sustaining Rice Systems for Latin America and the World	X	X
Tracking Light from the Sky Version 2.0 or Monitoring Rural Electrification from Space	X	X
"We feel fine": Big Data Observations of Citizen Sentiment about State Institutions and Social Inclusion	X	X
Open Roads Philippines: Improved Real Time Decision making of Infrastructure Investments for the Philippines by linking geospatial road network data with rich geo-tagged social data collected through mobile phones	X	X
Using High Resolution Satellite Imagery and Detection Algorithms to Better Track Poverty in Pakistan	X	X
Combining Taxi GPS Data and Open-Source Software for Evidence-Based Traffic Management and Planning	X	X
Evaluation of Crime and Infrastructure using Bayesian Maximum Entropy and Risk Terrain Modeling approaches in Bogotá, Colombia, 2008 - 2013	X	X
Big Data for User-focused Identification of Road Infrastructure Condition and Safety Concerns	X	X
The Sensors are Here! A High-Resolution Application on Understanding Individual Travel Patterns in African Cities	X	X

82 ●●● 여성가족 빅데이터 활용을 위한 기초연구

사업명	젠더 관점 여부	성 분리 여부
Using Sophisticated Content Analytics to Learn from World Bank Project Documents	X	X
Using Big Data to Predict Student Achievement in Low-Income School Settings	X	X
Big Data and the Cloud – Piloting “eHealth” for Community Reporting of Community Performance-Based Financing in Ghana	X	X
What’s for Dinner?	X	X
Leveraging the private sector (e-hailing service providers) through innovative partnerships to access and utilize taxi GPS data	X	X
Big Data for Transport Project, Morocco (BDT)	X	X
Using Geo-spatial Data to Secure Property Rights	X	X
Enabling up-to-date and accurate authoritative country mapping with crowdsourced geospatial data	X	X
Mapping the Risk Utility Landscape of Mobile Data for Development & Humanitarian Action	X	X
Data Visualization & Interactive Mapping to Support Response to Disease Outbreak	X	X
Understanding Public Perceptions of Immunization Using Social Media	X	X
Understanding Immunization Awareness and Sentiment through Analysis of Social Media and News Content	X	X
Analyzing Attitudes Towards Contraception & Teenage Pregnancy Using Social Data	○	○
Analyzing Social Media Conversations to Understand Public Perceptions of Sanitation	○	○
Using Twitter Data to Analyze Sentiment on Fuel Subsidy Policy Reform in El Salvador	X	X
Estimating Migration Flows Using Online Search Data	X	X
Mining Citizen Feedback Data for Enhanced Local Government Decision-making	X	X
Supporting the Post2015 Development Agenda Consultations Using Report	X	X
Using Twitter to Understand the Post2015 Global Conversation	X	X
Feasibility Study: Supporting Forest and Peat Fire Management Using Social Media	X	X
Using Twitter to Measure Global Engagement on Climate Change	X	X
Feasibility Study: Identifying Trends in Discrimination Against Women in the Workplace in Social Media	○	○
Feasibility Study: Analyzing Large-scale News Media Content for Early Warning of Conflict	X	X

사업명	젠더 관점 여부	성 분리 여부
Using Mobile Phone Activity For Disaster Management During Floods	X	X
Feasibility Study: Crowdsourcing High Frequency Food Price Data in Rural Indonesia	X	X
Analyzing Seasonal Mobility Patterns Using Mobile Phone Data	X	X
Using Mobile Phone Data and Airtime Credit Purchases to Estimate Food Security	X	X
Using Big Data Analytics in response to infectious disease outbreaks	X	X
Halting biodiversity loss making spatial data accessible to policy makers with Big Data tool	X	X
Analyzing Social Media Conversations to Understand Public Perceptions of Sanitation	○	○



Abstract

A Basic Study for Utilizing Gender Big Data

You Kyung Moon

Young Taek Kim

Geon Pyo Park

Chiyoun Cha

The importance of data is increasing from collection to its utilization day by day, and decision-making based on data is sine qua non for decision-makers and businesses as well as researchers. Big data, in particular, has become an essential concept in the use and analysis of data. This study defines the characteristics of Gender big data, and proposes future works for gender policy and its supporting researches.

This study examines which types of data are more likely to be utilized, and defines three characteristics based on it: firstly, data which have been produced so far for the gender policy; secondly, data which can be generated through the inclusion of gender perspective in the production and utilization of big data; thirdly, data that the established big data are sorted by sex or data reflecting themes related to women and family.

This study proposes to establish government plans for national strategy and data governance in favour of Gender big data. Especially, it is necessary (1) to sort raw data of women and family policy in big data; (2) to establish and to manage data-base of Gender big data; (3) to improve gender perspective in the government projects for big data; (4) to support fundamental researches for Gender big data.

The research areas which need to be studied in future are: (1) to excavate themes for women and family policy which needs big data research, and to examine the big data production plan; (2) to establish development plans on survey statistics in state statistics related to women and family policy; (3) to investigate on ways to increase the response rate in surveys through the link to big data.



2016 연구보고서(수시과제)-13

여성가족 빅데이터 활용을 위한 기초연구

2016년 11월 28일 인쇄

2016년 11월 30일 발행

발행인 : 이 명 선

발행처 : 한국여성정책연구원

서울특별시 은평구 진흥로 225

전화 / 02-3156-7000 (代)

인쇄처 : 도서출판 한학문화

전화 / 02-313-7593 (代)

ISBN 978-89-8491-833-7 93330

<정가 9,000 원>