

# **빅데이터 활용 전략 연구(미국)**

2020.5

**김 서 곤**

## 국외훈련 개요

1. 훈련국 : 미국
2. 훈련기관 : 켄트주립대학교  
(Kent State University, 미국 오하이오주)
3. 훈련과제 : 이동안민의 열린시정 실현을 위한  
빅데이터 활용 전략 연구
4. 훈련기간 : 2018. 8. 1.~2020. 6. 5.

# 차 례

## I. 서 론

## II. 빅데이터와 정부

1. 빅데이터의 정의
2. 빅데이터의 기원과 과정
3. 빅데이터가 공공 부문에 미치는 영향
4. 시민을 위한 빅데이터
5. 정부의 빅데이터 구현 과제
  - 가. 기술 관점
  - 나. 인적 관점
  - 다. 비즈니스 프로세스 관점

## III. 빅데이터와 정부 거버넌스

1. 빅데이터와 국가 전략
  - 가. 빅데이터와 국가 전략
  - 나. 국가별 빅데이터의 개방과 활용
2. 정부 거버넌스에 빅데이터 적용
  - 가. 빅데이터 적용 유형
  - 나. 빅데이터 처리 기술 및 응용

- 다. 빅데이터의 분석 기술 단계
- 라. 정부 거버넌스의 빅데이터 기준
- 3. 정부 거버넌스에 대한 빅데이터의 기회, 문제점, 대책 및 연구 과제
  - 가. 정부 거버넌스에 대한 빅데이터 기회
  - 나. 정부 거버넌스에 대한 빅데이터의 문제점
  - 다. 정부 거버넌스에 대한 빅데이터 대책
  - 라. 연구 과제

#### **IV. 미국 정부의 빅데이터 활용**

- 1. 미국 연방 정부의 빅데이터
  - 가. 용어의 사용
  - 나. 빅데이터 이니셔티브
- 2. 빅데이터와 정부의 변화
  - 가. 빅데이터 활용
  - 나. 빅데이터 소비
  - 다. 효율적이고 효과적 서비스
  - 라. 빅데이터 장벽 극복
- 3. 미국 주정부의 무결성 평가 사례 연구
  - 가. 정부 무결성 평가의 이해
  - 나. 빅데이터를 기반으로 한 정부의 무결성 평가

#### **V. 결 론**

# I . 서 론

빅데이터는 그 부분들의 합보다 오히려 더 많습니다. 그것은 규모에 관한 것이지만, 또한 이중 데이터 세트를 상호 연결하고 이를 마이닝하고 분석하는 방법에 관한 것이기 때문입니다. 빅데이터는 금융 시장에서 슈퍼마켓에 이르기까지 모든 곳에서 중요한 영향을 미치기 시작하였습니다. 앤디 윌리엄슨(Andy Williamson)이 설명한 것처럼, 빅데이터는 공공 정책 개발과 공공 서비스 제공 및 분석에 중요한 요소가 되기 시작하였습니다. 빅데이터는 세상을 둘러볼 수 있는 강력하고 새로운 방법을 제공하지만 동시에 소유권, 개인 정보 보호 및 오용의 문제를 초래하고 있습니다. 우리는 빅데이터의 채택이 증가함에 따라 법률과 관행이 보조를 맞추고 실수와 오용을 방지할 수 있도록 보장해야만 합니다.

빅데이터는 정부, 조직 및 시민이 다양한 목적과 목표를 특별한 방식으로 달성할 수 있도록 합니다. 공공 부문의 서비스 제공을 향상시키고 내부 및 외부 기관과의 교류를 위한 정부의 투자는 공공 서비스 투명성, 협업, 전자 참여 그리고 효율성을 향상시킬 것입니다. 현재 사용 가능한 방대한 양의 데이터는 조직에서 처리 및 분석하기가 어렵습니다. 따라서 오늘날 정부가 직면한 주요 과제 중 하나는 빅데이터의 구현이 될 것입니다. 그럼에도 불구하고 정부는 이니셔티브를 취하기 전에 효과적인 계획과 전략을 세우고 빅데이터 분석하기 위한 프로젝트를 효율적으로 채택하고 프로젝트에 대한 투자가 국가에 이익이 될 수 있도록 하여야 합니다.

정부 거버넌스에 빅데이터를 적용하는 것은 초기 단계에 있습니다. 앞으로 정부는 데이터 활용을 개선하고 부서와 국가간 격차를 좁히기 위해 세부 법률과 규칙을 수립해야 할 것입니다. 정부 관리자 및 직원의 능력과 열정을 향상시키는 메커니즘의 확립도 필요할 것입니다.

## II . 빅데이터와 정부

### 1. 빅데이터의 정의

빅데이터는 우리가 이해하는 것보다 훨씬 다양한 의미로 사용되는 까다로운 새로운 용어 중 하나입니다. 그것은 디지털 영역에서 나오는 수많은 새로운 용어들과 같이 논쟁의 여지가 있는 담론입니다: 유동적이고, 사용에 따라 진화하고 변화하며, 재사용되고 더욱 폭넓게 채택됩니다. 이름에 내재된 것은 규모의 개념이며, 빅데이터는 우리들의 쇼핑 목록보다도 훨씬 더 많습니다. 그러나 빅데이터는 단순히 데이터의 규모가 아니라 상호 연결의 규모, 즉 대규모 데이터와 이종 데이터 세트간에 존재하는 관계의 규모입니다. 따라서 빅데이터는 그 부분들의 합보다 큰 디지털 형상을 만들기 위해 서로 연결된 데이터인 것입니다. 데이터는 수동적으로 존재하기 때문에 우리는 그것의 사용여부와 다른 데이터와의 연결 여부를 결정할 수 있습니다. 빅데이터는 시장, 조직 및 시민과 정부 간의 관계 등을 변화시키는 방식으로 새로운 통찰력을 발굴하거나 새로운 형태의 가치를 창출하기 위하여 더 작은 규모로는 수행할 수 없는 대규모로만 수행할 수 있는 것들을 말합니다. 다시 말해, 차이를 만들어내는 단순히 큰 데이터가 아니라 그것을 가지고 우리가 무엇을 하느냐가 매우 중요한 것입니다.

분야별 빅데이터의 개념

분 야	빅데이터 개념	연구자
경영	인터넷 데이터, 모바일 트랜잭션, 사용자 생성 콘텐츠 및 소셜 미디어를 포함한 다수의 소스로부터 빅데이터가 생성되며 센서 네트워크, 영업 쿼리 및 구매 트랜잭션과 같은 비즈니스 트랜잭션을 통해 의도적으로 생성된 콘텐츠	(George, Haas, & Pentland, 2014)

분 야	빅데이터 개념	연구자
공공 정책	관리 데이터의 새로운 형식, 품질 및 가용성	(Pirog, 2014)
정치 과학	머신 러닝과 같은 기술 혁신을 통해 연구원들은 소셜 미디어 데이터와 같은 새로운 유형의 데이터 또는 적은 양의 기존 데이터를 적은 비용으로 수집 할 수 있었습니다.	(Clark & Golder, 2014)
정보 및 기술	크고 개방된 연결된 데이터(Big and Open Linked Data).	(Janssen et al., 2017)
전산 소셜 과학	관계의 구조와 내용에 대한 정보를 제공하는 장기간의 상호 작용에 대한 초 단위의 그림	(Lazer et al., 2009)

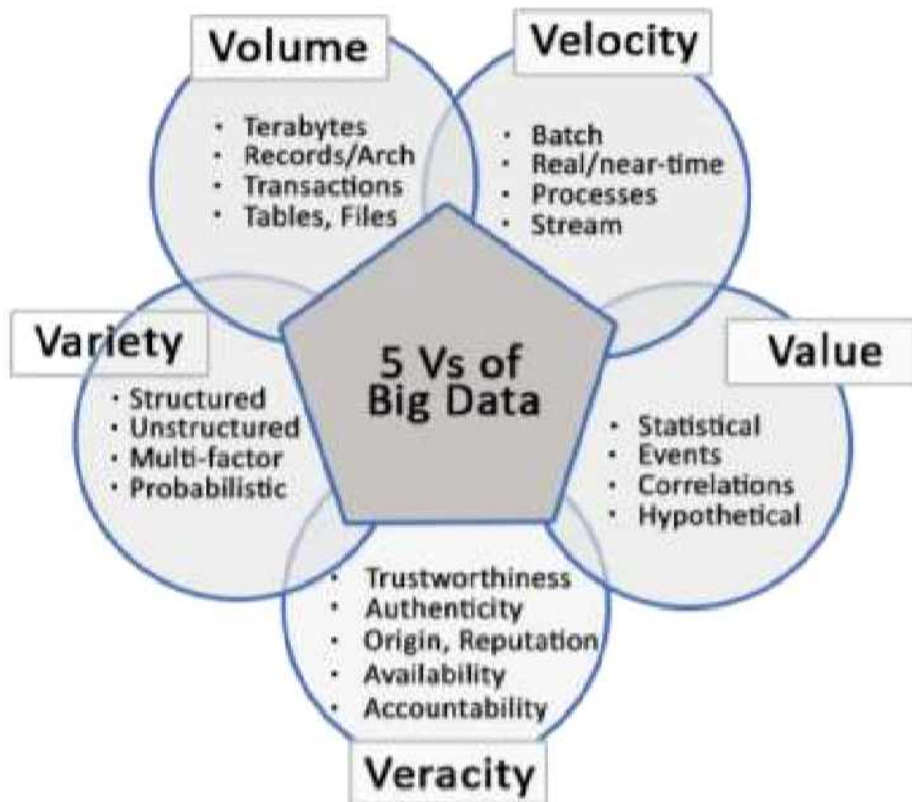
## 2. 빅데이터의 기원과 과정

대규모 데이터 세트는 오랫동안 우리와 함께 해 왔으며 은행 및 기타 대형 기관은 많은 양의 데이터를 오랫동안 보관해 왔습니다. 오늘날의 차이점은 저렴한 스토리지 및 컴퓨터 처리 능력으로 인해 더 많은 가능성을 제공하고, 더 스마트한 통계 및 계산 방법은 가치와 목적을 제공하며 데이터 집합을 연결하는 새로운 방법으로 의미와 문맥(context)을 제공한다는 것입니다. 따라서 빅데이터는 새로운 방식으로 저장, 처리 및 공유할 수 있는 대용량의 상호 연결된 데이터로, 해당 데이터가 나타내는 내용에 대한 풍부하고 심층적인 분석과 이미지를 제공합니다. 이를 통해 우리는 '데이터 마이닝' 및 '데이터 분석'의 개념을 이해할 수 있습니다. 여기서 강력한 새로운 컴퓨터 처리 기술을 사용하여 서로 연결된 방대한 데이터 세트를 발견, 처리 및 분석하여 패턴, 추세 및 정서와 같은

것들을 파악할 수 있습니다. 여기에는 엄청난 상업적 이점과 공공 부문 잠재력이 있으며 이러한 시장에서 멀리 어답터가 된 조직은 상당한 경쟁 우위를 나타낼 수 있습니다. Facebook은 이에 대한 좋은 예입니다. Facebook은 매일 상상할 수 없을 정도로 방대한 양의 데이터를 생성합니다. 그리고 이 데이터는 우리에게 광고를 타겟팅하고 사용자들에 대한 복잡한 이해를 위해 활용됩니다. 온라인 세계에서 비용이 제로이기 때문에 당신이 제품이라는 것은 비교적 명백한 사실입니다. 페이스북의 경우에는 마찬가지일 것입니다.

우리가 아는것처럼, 빅데이터의 적용은 기술적 관점을 뛰어 넘습니다. 그것은 프라이버시, 동의, 법률 체계 및 그것을 둘러싼 정치 문제뿐만 아니라 미지에 대한 자연적인 인간의 두려움과 이것이 만들어내는 내재된 불신의 이슈를 매우 빠르게 확산시킵니다. 인터넷 덕분에 우리가 만드는 데이터의 양은 전례없는 속도로 증가하고 있습니다. 과거에는 데이터가 주로 기업 및 공공 부문에서 만들어졌으며 비공개로 유지되었으며 내부적으로 사용되었습니다. 인터넷 기반 네트워크의 부상으로 인해 데이터를 더 많이 사용할 수 있게 되었으며 인터넷이 소셜웹으로 발전함에 따라 데이터 생산량이 크게 증가했습니다. 그리고 대부분은 Facebook, You Tube 및 Twitter와 같은 플랫폼에서 개인들이 생성합니다. 현재 약 6 zettabytes의 디지털 데이터가 있으며 매년 50% 정도 증가하고 있는 것으로 추정됩니다. IBM에 따르면 2012년에는 매일 25억 기가 바이트의 데이터가 생성되었습니다. 그들은 데이터의 약 75%가 구조화되어 있지 않다는 것을 알고 있습니다. 즉, 기존 데이터베이스에서 보유하고 있는 친숙한 종류의 구조적이고 독점적인 데이터가 아니라 텍스트, 음성 및 비디오와 같은 소스에서 가져온 것입니다. 이 모든 데이터 중에서 데이터의 34%만이 유용한 것으로 추정됩니다). 그러나 데이터의 7%만이 태그(tag) 되어 있으며 문맥과 의미를 제공하기 위해 존재하는 모든 데이터의 1% 만 실제로 분석되었습니다. 데이터가 빅데이터가 되기 위해서는 몇 가지 조건을 충족시켜야 합니다





빅데이터의 특징

○ Volume (양)

볼륨은 빅데이터의 필수 특성입니다. 많은 응용 프로그램에서 사용된 빅데이터 세트의 크기를 설명합니다. 볼륨은 데이터 측정과도 관련이 있습니다. 분석 결과 및 일반 스토리지에서 관리되는 데이터와 비교할 때 엄청난 정보 세트입니다. 빅데이터 세트의 크기는 여러 terabytes 및 petabytes로 나타냅니다. 또한 크기는 페타 바이트보다 클 수 있으며 exabytes 또는 zettabytes 일 수 있습니다. 크기도 다양한 소스에서 생성된 모든 유형의 데이터를 가리킬 수 있습니다. 더욱이, 데이터, 트랜잭션, 다수의 테이블 및 다수의 파일의 보고서를 계산함으로써 빅데이터를 정량화 할 수도 있습니다.

○ Variety (다양성)

다양성은 다양한 데이터 세트 및 소스입니다. 내부 또는 외부 리소스와 같은 여러 리소스에서 데이터가 생성될 수 있습니다. 다양한 리소스에서의 데

이터 기록으로 인해서 데이터 세트에 많은 변화들이 생기게 됩니다. 다양성은 데이터 세트에서 구조적 일관성이라고도 합니다. 데이터는 세 가지 클래스로 분류 할 수 있습니다. (i) 구조화 - Structured (ii) 반 구조화 - Semi Structured (iii) 비 구조화 - Unstructured 데이터. 사진, 오디오, 비디오 및 기타 리소스에서 생성되는 데이터는 분류하기가 어렵습니다. 일반적으로 텍스트, 이미지, 비디오, 오디오는 구조화되지 않은 데이터 세트의 일반적인 예이며, 분석을 위해 가상 머신에 필수적인 구조적 배열은 거의 없습니다. 또한 XML(Extensible Markup Language)은 반 구조화된 데이터의 일반적인 예입니다. 반 구조화된 데이터는 고정 필드에 해당하지 않지만 데이터 요소를 배포하기 위한 태그를 보유하고 있습니다. 다양성은 데이터가 소스의 모든 측면에서 가져 온다는 것을 의미합니다. 즉, SQL과 같은 전통적 데이터베이스일 수 있는 구조화된 데이터, 반 구조화된 그러나 데이터베이스와 같은 공식 또는 통합 구조가 없는 데이터 그리고 조직화되지 않은 데이터를 의미하지만 비즈니스 인텔리전스가 없는 비 구조적 데이터 소스들이 있습니다.

#### ○ Velocity (속도)

속도는 데이터가 생성되는 속도와 데이터를 분석하고 설명하는 속도로 소개될 수 있으며 데이터 생성 횟수 또는 데이터 전송 빈도로도 정의 할 수 있습니다. 일반적으로 속도는 생산 속도이며 빅데이터의 경우 상당히 높습니다. 데이터가 크게 증가하면 데이터를 더 빨리 해석해야 함을 의미합니다. 즉, 데이터가 빠르게 증가할수록 데이터에 대한 요구사항들은 더욱 더 커집니다. 궁극적으로 속도는 전달, 생성 및 처리되는 데이터의 속도로 설명 할 수 있습니다.

#### ○ Veracity (정확성)

정확성은 데이터 불일치, 모호성, 불완전성, 지연시간, 근사 및 속임수로 인해 빅데이터 품질의 특성이 나빠거나 양호하거나 정의되지 않은 특성 또는 특징을 말합니다. 이것은 빅데이터의 품질에 중점을 둡니다. 일반적으로 일부 데이터의 예측 불가능성과 관련된 정확성은 신뢰할 수 있는 예측을 얻기 위해 빅데이터 영역에서 분석이 필요합니다. 또한 데이터의 정확성

(accuracy)으로 정의될 수 있으며, 데이터는 실제 다양한 리소스들에서 수집되고 데이터의 보안이 제공되어야 합니다.

#### ○ Value (가치)

가치는 빅데이터의 속성으로 정의되며 Oracle에 의해 소개되었습니다. 빅데이터는 일반적으로 대략 "낮은 값의 밀도"를 특징으로 합니다. 따라서 원래 형식으로 얻은 데이터는 일반적으로 크기 또는 볼륨과 관련된 가치가 낮습니다. 그러나 정확한 크기의 큰 데이터를 분석하면 높은 가치를 얻을 수 있습니다. 또 다른 가치는 조직에 비즈니스 가치를 제공하는 것이나 조직 목표와 비즈니스 가치를 함께 결합하는 것은 어려운 작업입니다. 결론적으로, 빅데이터는 조직을 발전시키기 위하여 부가 가치를 창출해야 합니다.

### 3. 빅데이터가 공공 부문에 미치는 영향

빅데이터가 인터넷 혁명의 산물이라면 공공 서비스 제공에 어느 부분이 적합한지 이해하기 위해서는 두 가지 각도에서 주제에 접근해야 합니다. 첫째, 기술 채택은 사회 전체에서 바로 이루어 지므로 정부가 새로운 기술의 혁신가 또는 얼리 어답터가 아닐 수 있지만 결국에는 이러한 새로운 도구 중 일부를 채택하게 될 것입니다. 그러나 디지털은 고립되어 있지 않으며, 우리 사회는 기술적으로 결정적이지 않으며 정부의 빅데이터에 대한 맥락이 문화적이라는 것을 인식하는 것이 중요합니다. 지난 30년 동안 선진국의 많은 정부는 프로세스 개선과 서비스 제공에 의해 몰두된 기술 엘리트로 변모했습니다. 이러한 정부 시스템은 더욱더 조직화 되어 있고, 일부는 개방적이나 대중을 보호하기 위한 법률에도 불구하고 불투명한 경우가 자주 발생합니다. 최근에는 영국에서 목격된 긴축(austerity) 문화로 인해 정부가 더욱 약해졌습니다. 효율성에는 어떤 원동력이 있으며 디지털 플랫폼은 그 효율성의 중요한 동인으로 여겨집니다. 영국의 공공 부문 빅데이터도 기본적으로 디지털이라는 맥락에서 존재하므로 공공 서비스 제공은 디지털 전달 및 사용에 최적화되어 있습니다. 이미 빅데이터가 반드시 공개적인 것은 아니라고 말했지만 정부 상황에서는 그럴 가능성이 높습니다. 따라서 공공 부문의 투명성과 데이터 공개에 대한 증가하는 요구에 대한 맥락을

간략히 살펴볼 가치가 있습니다. 이것은 디지털 정부 의제 내에서 눈에 띄는 역점 사항이었습니다. 이는 민주주의를 디지털화하는 데 있어 한 가지 측면일 뿐이지만, 데이터는 더 나은 의사 결정을 내리고 시민의 눈높이를 맞추는데 있어서 중요한 엔진인 것입니다. 오픈 공공 데이터는 보다 적극적이고 더 나은 민주주의를 가능하게 합니다.

정부가 투명성을 향해 나아가고 개방된 정부 원칙을 채택하는 데는 여러 가지 이유가 있으며, 시민들이 기본적으로 데이터를 공개하도록 요구해야 하는 많은 이유가 있습니다. G8 정부가 이미 이를 정확히 이행하기 위해 노력했지만, 이행이 의도된 것보다 많이 뒤쳐져 있습니다. 아마도 빅데이터의 강력한 동인은 긴축(austerity)일 것입니다. 연결, 분석 및 어플리케이션을 통해 정부 데이터 세트를 더욱 효과적으로 사용하는 것은 효율성 측면에서도 중요합니다. 정부 운영의 전반적인 효율성을 개선하고 사기와 오류를 줄이기 위한 노력을 가속화하며 실제 세금 징수와 이론적 부채 간의 세금 격차를 해결하는 등 많은 분야가 있습니다.

빅데이터 및 관련 기술을 사용하여 정부는 많은 잠재적인 비용 절감 효과를 기대할 수 있으며, 이 중 대부분이 발생하는 상황을 보다 잘 이해한 결과로 수행될 수 있는 운영 및 프로세스의 개선을 통해 이루어질 수 있습니다. 정부 기관은 빅데이터를 사용하여 정부가 자체 데이터 및 측정값들을 활용할 수 있을 뿐만 아니라 다른 데이터 소스의 광범위하고 풍부한 환경을 마이닝할 수 있습니다.

정부에 대한 사람들의 의견, 부정적인 의견과 불만을 제기하는 정책, 서비스 또는 제공 업체를 이해함으로써 실패한 학교, 병원 또는 계약자를 식별할 수 있도록 하여 소셜 미디어의 데이터를 자기 개선에 사용합니다. 그들은 자신의 사이트나 다른 기업들의 사이트를 통해 그러한 데이터를 요청할 수 있습니다. 또한 Google 검색 API 또는 Google 트렌드에서 사람들의 관심사나 찾고 있는 것들을 파악하여 인터넷 사용자의 검색 패턴을 기록할 수 있습니다.

그러나 만병 통치약은 아니며, 공공 정책 목적을 위해 데이터를 개방하고 적극적으로 연결하여 빅데이터를 만드는 데는 큰 어려움이 있습니다. 또한 개인 정보 보호 문제, 개인의 권리 보호 방법 및 익명 데이터 세트 측면에서 수용 가능한 확실성을 보장하기 위해 어떤 방법을 사용할 수 있는지는 무시할 수 없습니다. 이러한 문제들은 특히 정부가 기업에 헬스 또는 과세 데이터 판매를 제안

할 때 주의해야 할 부분입니다.

이러한 변화는 부분적으로 정부가 서비스 제공 및 거래 수준에서 마이닝 및 분석 측면에서 데이터를 보다 똑똑하게 사용하기 때문에 발생하지만 정책 수준에서도 이점이 있습니다. 그러나 이런 일이 발생하기 위해서는 앞으로 공무원과 대중이 빅데이터를 사용하고 이해하는 데 있어 더 효과적으로 판단해야 하는 것이 중요합니다.

폭발적인 데이터와 데이터 조작 능력은 인구 규모에 가까운 사람들의 삶에 대한 통찰력을 제공합니다. 이것은 인간 계층 매핑이 의학에 영향을 준 것처럼 근본적으로 사회 정책을 바꿀 수 있습니다.

요약하면 공공 부문의 빅데이터 이점은 다음과 같습니다.

- 현재 이질적인 정부 및 공공 기관에서 데이터를 공유하지만 엄격한 통제 및 모니터링이 적용됩니다.
- 새로운 기술, 새로운 종류의 정보 및 관련분야 지식이 필요하며 그것이 복잡성을 증가시킬 수 있음을 인식하고 데이터의 가용성과 상호 연결성 및 새로운 도구 사용을 통해 학습합니다.
- 실시간 데이터 연결 및 세분화된 기능을 통해 개인화가 가능해졌습니다. 그 반대의 경우 기밀성 및 익명성에 대한 우려가 있으므로 데이터를 신중하게 제어하고 적절한 경우 식별 기능없이 집계해야 합니다.
- 빅데이터 세트를 함께 연결하면 복잡한 문제를 해결하고, 예측 분석 및 건강 기록을 분석하여 인근 범죄 또는 약국 재고의 패턴을 예측하는 등의 예측 분석 기회를 제공할 수 있습니다. 그러나 우리는 이것이 개인적인 이야기와 사람들의 생생한 경험을 희생하여 이루어지지 않도록 해야 합니다.
- 마지막으로, 빅데이터는 공공 부문과 민간 부문 모두에서 보다 광범위한 디지털 및 혁신 경제의 일환으로 혁신을 가능하게 하기 때문에 정부에 유용한 도구입니다.

#### 4. 시민을 위한 빅데이터

빅데이터는 정부가 공공 서비스를 개인화하여 보다 표적화되고 효율적이며 효과

적으로 만들 수 있는 기회를 제공합니다. 반대로 오용, 대량 감시, 개인 정보 보호 위반 및 시민의 과도한 상태 모니터링 기회를 제공합니다.

빅데이터는 정부가 공공 서비스를 관리하는 방법을 향상시키고 완전히 새로운 종류의 가치를 창출할 수 있게 합니다. 그러나 빅데이터 틀은 의심할 여지없이 정부 권력이 억제되지 않고 발생할 가능성을 증가시킵니다.

디지털 기술은 방해자 또는 작동자 역할을 할 수 있습니다. 빅데이터는 테러를 저지하거나 독감의 영향을 완화하는 데 사용될 수 있지만, 이는 국가가 사람들의 활동을 추적하고 모니터링 해야하는 절충안과 함께 발생합니다. 빅데이터는 분명한 경제적 이점이 있지만 민주적인 이점도 있기 때문에 공개 데이터를 공개하고 액세스할 수 있어야 합니다. 개방형 데이터는 책임을 제공합니다. 보다 광범위한 공적 투명성 철학의 일부로서, 정부가 정보를 공유하고 이제는 개인들이 보유할 수 있는, 그리고 그것을 용이하게 만드는 기술의 이용가능성이 증가하고 있습니다. 여기서 중요한 문제는 신뢰입니다. 투명성과 신뢰는 본질적으로 연결되어 있습니다. 불행하게도 우리 중 일부는 정부를 신뢰하지 않습니다. 대부분의 국민이 정부가 소규모 엘리트의 이익을 위해 행동하고 있다고 믿고 있으며, 더러운 정치에 노출된 것으로 추정하고 있습니다. 이 가정은 근거가 없는 것이 아니며, 몇몇 나라에 국한되지 않습니다. 더 나은 품질의 정보에 더 잘 액세스하면 민주적 프로세스에 대한 대중의 신뢰 재건을 지원할 수 있습니다. 개인 및 사회 차원에서 빅데이터에 대한 공개 액세스는 대중의 선택을 증가시킵니다. 정부는 시민이 직접 지불, 개인 예산, 자격 또는 선택을 통해 어떻게 자신의 삶을 통제할 수 있는지 설명합니다. 빅데이터를 사용하여 비교 정보를 제공하고 공공 서비스에서 의미있는 선택을 할 수 있습니다. 이에 대한 예로는 서비스 품질을 비교 분석하고 보다 폭 넓은 공개 토론과 효과적인 사용자 참여를 지원하여 정책에 더 잘 반영하고 서비스 디자인을 개선할 수 있습니다. 우리는 시민들이 빅데이터를 이용할 수 있게 함으로써 정부와 교류할 때 더 많은 정보와 자원들을 얻을 수 있는 이점을 고려해야 합니다.

이미 빅데이터의 잠재적 이점은 위험에 의해 어느 정도 상쇄된다고 언급했습니다. 빅데이터에는 강력한 거버넌스가 필요하며 프로세스와 표준이 없기 때문에 그 정확성이 떨어질 수 있습니다. 빅데이터에서 발생하는 강력한 위험은 개인 정보 보호 및 관련 대중 인식 문제입니다. 대중의 이해는 일반적으로 낮으며 공개 데이터는 일상 생활에 분명한 이점이 없는 추상적인 문제로 보였질수 있습니다

다. 즉, 공개 데이터의 의미가 정확히 무엇인지에 대한 대중의 명확성이 부족할 수 있습니다. 특히, 시민들이 개인 및 공공 데이터의 개념과 충돌하는 것처럼 보일 수 있습니다. 또한 개인 정보 문제는 개방형 및 연결된 데이터에 대한 많은 사람들의 우려의 핵심입니다. 정부 정책에 명확하게 표현되어 오픈되어 있는 모든 데이터를 익명으로 처리해야 합니다. 공개 시험 데이터 평가 및 투명성 패널 작성을 포함한 모범 사례 개발을 통해 프라이버시 요구 사항을 해결해야 하며 기술 및 개인 정보의 법적 측면을 다루어야 하며, 데이터가 개인 정보에 영향을 미치는 경우 적절한 통제를 고려해야 합니다.

## 5. 정부의 빅데이터 구현 과제

정부에서 빅데이터를 구현하려는 여정에는 직면한 많은 과제가 있습니다. 정부 이니셔티브는 이러한 새로운 트렌드의 시행과 관리에 있어서 광범위한 해결 과제와 방해물의 존재와 부딪히고 있습니다. 정부가 빅데이터를 구현하고 이러한 과제에 대해 가능한 솔루션 그리고 빅데이터 시대를 통해 정부 서비스를 통합하는 데 있어 주요 과제들을 3가지 제안 영역으로 분류할 수 있습니다. 첫째는 기술 관점에서, 둘째는 사람 관점에서, 세 번째는 비즈니스 관점으로 다음과 같습니다.

### 가. 기술 관점

빅데이터 분석에는 새로운 기술, 프로세스 최적화 기술, Hadoop, Spark 등의 빅데이터 기술을 사용하는 데이터 관리 도구를 사용하여 데이터를 저장 및 분석하는 기능이 필요합니다. 볼륨 증가에 따라 추가 데이터 스토리지 시스템, 스토리지 메커니즘, 대규모 데이터 수요를 충족시키는 새로운 환경과 기술이 필요합니다.

이 새로운 혁명에는 빅데이터를 의미있는 부가 가치로 옮기기 위한 효율적인 프로세스가 필요합니다. ICT 인프라가 충분하지 않으면 정부의 빅데이터 응용 프로그램을 수행하기가 어렵습니다. 빅데이터는 모든 유형의 형식, 기존 데이터베이스의 데이터 구조, 단순한 구조 또는 전자 메일 및 금융 거래와 같은 비정형 데이터 형식으로 제공되므로 큰 처리 능력이 필요합니다.

또한 매일 증가하는 데이터의 양과 급속한 증가는 동일한 속도로 데이터를 모델링하고 분석하는 능력과 용량보다 더 앞서잡니다. 기술 관점에서 다음과 같이 과제를 요약할 수 있습니다.

- IT 및 인프라(기술 인식, 전자 준비(e-readiness), 컴퓨터 활용 능력, 통신 장비)의 능력.
- 데이터 보안 문제 및 정책 문제(법률 및 규제).
- 빅데이터 분석 기술과 같은 빅데이터의 인적 자본에 대한 전문가 및 기술 부족
- 빅데이터 조직의 통제 부족.
- 기존 스토리지 시스템과의 호환성 부족으로 현재 스토리지 시스템은 이러한 방대한 양의 데이터를 지원할 없음.
- 빅데이터의 급속한 성장은 이러한 데이터의 모델링과 분석에 선행.
- 빅데이터 기술을 사용하는 데이터 관리 도구(Hadoop, Spark 등)의 부족.

## 나. 인적 관점

온라인 서비스 제공 업체는 고객이 입력, 탐색 및 클릭 한 모든 데이터를 저장하고 관리합니다. 서비스 제공 업체는 고객이 누구인지, 무엇을 하는지, 거주지, 선호하는 것을 알 수 있습니다. 또한 온라인 서비스 제공 업체는 사용자의 데이터를 다른 제 3자 및 광고주에게 판매할 수 있으므로 그들은 원하는 정보를 사람들에게 광고할 수 있습니다. 따라서 사람들은 공유할 수 있는 것과 공유할 수 없는 것에 대해 교육을 받아야 합니다. 주요 문제는 개인 정보를 모두 보호할 수 없으며 빅데이터를 관리하는 제 3자가 모든 소셜 네트워킹 활동을 읽을 수 있다는 것입니다. 불행히도 많은 사람들이 회사가 빅데이터를 어떻게 사용하고 있는지에 대해 거의 이해하지 못합니다. 인적 관점에서 다음과 같이 과제를 요약할 수 있습니다.

- 인적 자본 개발 및 학습 기술, 능력 및 경험 부족.
- 문화 저항력.
- 사람들은 기술을 믿을 수 없음.



## 다. 비즈니스 프로세스 관점

대량의 데이터가 공공 서비스에 가치를 창출하기 때문에 빅데이터는 정부 서비스를 확대할 것입니다. 정부의 지원이 필요하며 정부는 빅데이터 구현과 관련하여 더 많은 연구와 파트너십을 지원해야 합니다. 정부가 지식 경제를 달성하고 경쟁 우위와 부가가치를 강화하는 도구로 전자적 참여를 수행할 때 빅데이터는 지식을 창출하고 경쟁력을 자극하며 성과를 향상시키며 효과적인 의사 결정을 제공할 수 있습니다. 빅데이터에 대한 투자는 비즈니스 프로세스 관점에서 빅데이터에서 고품질의 가치를 얻기 위해 해결해야 할 중요한 과제를 제기하고 다음과 같이 요약할 수 있습니다.

- 사업 전략 변경(비전, 임무, 목적, 목표)
- 변형과 관리 변화(구조, 문화, 기술, 과정, 변화에 대한 저항).
- 파트너십 및 협업, 개인 및 민간 파트너십.
- 커뮤니티 및 네트워크 구축.
- 리더십 역할(동기 부여, 참여, 영향력, 지원).

정부에서 빅데이터를 구현하는 데 있어 3가지 주요 과제에 대한 가능한 솔루션을 다음과 같이 제안할 수 있습니다.

과 제	해결책
기술 관점	<ul style="list-style-type: none"> <li>○ 스토리지 용량을 늘리고 클라우드 스토리지와 같은 사용 가능한 스토리지를 활용하여 빅데이터를 저장 및 처리.</li> <li>○ 보안 문제, 정책 문제, 법률 및 규정 구현.</li> <li>○ 효과적인 데이터 관리 도구를 사용하여 여러 소스의 구조, 비 구조 및 반 구조 데이터를 분석.</li> <li>○ 효율적인 빅데이터 리소스 제공 및 통합.</li> </ul>

과 제	해결책
인적 관점	<ul style="list-style-type: none"> <li>○ 데이터 분석, 빅데이터 보안과 같은 빅데이터의 가용 인적 자본의 기술 향상 및 빅데이터의 구현을 지원할 수 있는 전문가 수급.</li> <li>○ 빅데이터와 정부 간의 협업을 지원하는 국가 이니셔티브 구현.</li> <li>○ 빅데이터와 자치 정부에 미치는 영향에 대한 인식 증대.</li> <li>○ 시민들이보다 창의적이고 소셜 네트워크를 효과적으로 표현할 수 있도록 역량강화.</li> </ul>
비즈니스 프로세스 관점	<ul style="list-style-type: none"> <li>○ 정부에서 빅데이터를 구현하기 위한 로드맵을 지원하는 효과적인 전략을 수립.</li> <li>○ 공공 및 민간 파트너십 구축.</li> <li>○ 빅데이터의 역량 아래 지역 사회와 정부 사이의 네트워크 구축.</li> <li>○ 빅데이터 환경으로 연결되는 효과적인 전략과 로드맵을 수립.</li> <li>○ 리더십은 정부에서 빅데이터 구현에 동기를 부여하고 참여하고 영향을 미치고 지원.</li> <li>○ 정부의 중요한 빅데이터를 의사 결정 프로세스에 통합.</li> <li>○ 전략적 계획과 의사 결정을 구성하는 데 있어 리더십 및 정보 관리의 역할을 강화.</li> </ul>

### Ⅲ. 빅데이터와 정부 거버넌스

최근에는 사회 거버넌스, 정부 거버넌스, 국가 거버넌스에서 빅데이터가 핫스팟이 되었습니다. 빅데이터는 사회 보장, 도시 건설, 고용 및 기업가 정신, 환경 및 지속 가능한 개발과 같은 문제를 해결하는 데 사용될 수 있습니다. 빅데이터는 부패의 범칙을 밝히는 데 사용될 수 있습니다. 빅데이터를 통해 도시 거버넌스에 대한 방대한 정보를 수집하여 도시의 올바른 거버넌스를 달성할 수 있습니다. 빅데이터에 대한 이론적 연구가 본격적으로 진행되었지만 빅데이터와 공공 관리의 조합은 이제 막 시작되었습니다.

#### 1. 빅데이터와 국가 전략

세계의 선진국은 빅데이터에 큰 중요성을 부여하고 빅데이터 국가 전략을 제시합니다. 여기서는 미국, 영국, 호주, 프랑스 등의 데이터 국가 전략을 비교하여 보았습니다.

##### 가. 빅데이터와 국가 전략

첫째, 국가별 빅데이터) 2012년 3월, 백악관 과학 기술 정책국(OS-TP)은 "빅데이터 연구 및 개발 프로그램"을 주도했으며 "미국 연방 정부의 지침을 통해 빅데이터"를 공개했습니다]. 2013년 8월 호주 정부는 "호주 공공 서비스 정보 및 통신 기술 전략 20122015"을 발표했습니다. 2013년 10월 영국은 "수익 창출 기회 : 영국 데이터 기능 전략"을 발표했습니다. 프랑스는 "2015 년 디지털 공화국을 촉진하기 위한 활동 계획"을 시작하였습니다. 2015년 8월 31일, 중국 국무원은 "빅데이터 개발 촉진을 위한 행동 계획 개발에 관한 통지"를 발표하여 중국의 빅데이터를 국가 전략에 표시했습니다.

둘째, 국가별 빅데이터 목표) 미국에서 빅데이터의 목표는 최첨단 핵심 기술을 수집, 저장, 유지, 관리, 분석, 개발 및 공유하는 것입니다. 과학 기술 탐

색 속도를 높이고 국가 안보를 강화하고 기존 관리를 변경하기 위해 핵심 기술을 사용합니다. 호주 공공 서비스 빅데이터 전략의 주요 목표는 ICT를 계속 사용하여 정부 역량을 향상시키고, 사람, 지역 사회 및 비즈니스 상호 작용을 향상 시키며, 언제 어디서나 사람들이 필요할 때 더 효율적이고 편리한 정부 서비스를 제공하는 것입니다. 영국 정보 경제위원회의 목표는 국가가 데이터 혁명을 통해 21세기의 리더가 되도록 하는 것입니다. 국가는 점점 증가하는 데이터 가치를 포착하고 부와 사회적 이익을 창출하며 삶을 변화시킬 수 있습니다. 프랑스 디지털위원회의 목표는 프랑스 국민으로부터 데이터, 의견 및 제안을 수집하고 다음 행동 계획 및 결정을 공식화하고 개발하는 것입니다. 중국의 빅데이터의 목표는 법률 및 규정의 건설 속도를 높이고 시장 개발 메커니즘을 개선하며 표준 시스템을 확립하며 재무 및 재정 지원을 강화하며 전문 인력 교육을 강화하며 국제 교류 및 협력을 촉진하는 것입니다.

## 나. 국가별 빅데이터의 개방과 활용

첫째, 정부마다 다른 정도의 데이터 개방성을 가진다) 버락 오바마 미국 대통령은 투명하고 개방된 정부 계획에 서명하고 Data.gov를 발기하였습니다. 2010년 7월, 호주 정부는 개방형 정부를 세우기 위해 2.0태스크 포스 보고서의 핵심 권고안을 발표했습니다. 2010년 영국은 빅데이터 인프라 구축을 지원하기 위해 1억 9,900만 유로를 투자하고 전자 인프라 리더십위원회(E-Infrastructure Leadership Council)를 설립했으며 data.gov.uk에 10,000 개 이상의 공공 데이터베이스를 게시했습니다. 2015년 프랑스는 www.data.gouv.fr를 통해 French Open Public Data Platform을 설립했습니다. 중국은 최근 베이징, 상하이, 충칭, 광저우, 구이양 및 기타 지방과 도시는 정부 데이터를 열고 공유할 수 있는 빅데이터 서비스 플랫폼을 구축했습니다.

둘째, 정부마다 데이터 사용량이 다르다) 맥킨지 인스티튜트 2016 (McKinsey Institute 2016)은 정부와 기타 분야에서 데이터 사용에 큰 차이가 있다고 보고했습니다. 정보 통신 기술 분야는 최전선, 미디어 및 재무 부서에 있으며, 여러 전통 산업이 마지막에 있습니다. 전반적으로 정부 부서의 빅데이터 사용량은 상대적으로 적으며 유럽에서는 12 %에 불과합니다

다. 정부 기관 간의 데이터 공유에는 기술 관리 기능이 중요하지만 미국 정부와 비정부 기구의 설득력 있는 메커니즘과 기술 참여 기능에 의해 결정되었습니다.

## 2. 정부 거버넌스에 빅데이터 적용

### 가. 빅데이터 적용 유형

정부 거버넌스에 빅데이터를 적용하는 것은 거시 통제 결정, 위험 경고 및 감독입니다.

첫째, 정부 거버넌스에서 빅데이터에 대한 거시적 통제 결정) 빅데이터를 사용하여 정부는 민사, 시장 감독, 사회 질서, 공공 서비스 등 인력 흐름에 대한 좋은 정보를 보유할 수 있으며, 데이터 마이닝 기술을 바탕으로 공중 보건 서비스 및 사회 서비스 관리 역량의 현대화를 촉진합니다. 연구원들은 Google 트렌드와 트위터의 빅데이터를 사용하여 미국 우주 정책 개발과 시민 간의 관계를 분석합니다. 보스턴 정부는 수집한 60 만 건 이상의 서비스 청구와 311 개의 빅데이터 시스템을 사용하여 객관적인 사건과 저장 기록 사이의 관계를 분석했습니다.

둘째, 정부 거버넌스에서 빅데이터의 위험 경고 시스템) 유럽위원회(European Commission)는 지능형 환경 스캐닝 레이더와 첨단 환경 스캐닝 프로세스를 사용하여 미래 범죄 모델을 결정하는 "e-POLICE"라는 프로젝트에 자금을 지원했습니다. 유럽법 집행 기관과 협력한 COURAGE 프로젝트는 사이버 범죄에 대한 유럽 연구 의제를 결정하여 사이버 공격 및 남용으로부터 유럽의 중요 인프라를 보호하는 방법을 결정하고 테러리스트가 추종자들을 홍보하고 채용하기 위해 인터넷 사용하는 것에 대처하는 것을 목표로 합니다. 정부의 의사 결정은 데이터 분석, 시각화, 세계 표준 테스트 기준, 전자 모니터링, 위키피디아 편집자, 샌프란시스코 주택 위기 및 신용 카드 데이터베이스를 기반으로 훌륭한 모델을 생성하고 사건을 예측할 것입니다. 시맨틱 정보 이미지를 통해 사람들의 행동을 자동으로 모니터링하고 신속하게 피드백하여 가능성 있는 불확실한 결과를 예측할 수 있습니다.

셋째, 정부 거버넌스에서 빅데이터의 집행 감독 시스템) 보스턴 마라톤 폭탄

테러의 경우 소셜 미디어는 대규모 글로벌 정보와 사진을 통해 용의자를 신속하게 식별했습니다. 트위터, 페이스북 및 기타 사이트가 차르나에프 형제를 식별하고 체포하는 데 도움이 된 것으로 생각됩니다. ATHENA 프로그램은 소셜 미디어, 특히 스마트 모바일 기기에 중점을 두어 대중이 먼저 도움을 필요로 하는 사람들을 도울 수 있도록 합니다. Odyssey 프로젝트는 총기 범죄의 통계 패턴(예 : 총기, 탄약 및 범죄 배경)이 어떻게 사용되는지 분석합니다. 법 집행 기관은 툴과 응용 프로그램을 사용하여 총 범죄 데이터를 백엔드 시스템과 통합하여 보다 많은 정보를 제공하는 설문 조사를 지원 합니다.

## 나. 빅데이터 처리 기술 및 응용

첫째, 빅데이터의 처리 기술은 배치 기반 처리 도구 및 스트림 기반 처리 도구입니다. 빅데이터, NLP, Hadoop, NoSQL의 클라우드 컴퓨팅은 반 구조화 및 비 구조화 정보 분석을 가능하게 합니다. 새로운 컴퓨팅 아키텍처뿐만 아니라 센서, 무선 주파수 장치(RFID) 및 기타 전자 감지 장치는 병렬 컴퓨팅 아키텍처를 형성하고 복잡한 지역, 산업 및 영역에서 데이터를 수집하여 복잡한 분석, 통합, 미적분 및 예측을 수행합니다.

둘째, 분석 기술을 정부 거버넌스에 적용) 데이터 마이닝 기술을 사용하여 TOP1000 온도 검색 및 가장 빠른 TOP1000 온도 검색을 수행하고 시맨틱 마이닝 기술을 사용하여 여론의 일반적인 상황과 구조를 분석할 수 있습니다. Hadoop 분산 파일 시스템을 통해 안전하지 않은 동작을 자동으로 모니터링하고 신속하게 피드백 하여 가능성 있는 불확실한 결과를 예측할 수 있습니다. R 디지털 컴퓨팅 프로그램 기반한 프로젝트 투표 소프트웨어 (pvs)는 대중 투표를 예측할 수 있습니다. 빅데이터 플랫폼에 내장된 지식 분석 및 임시 예측 모델을 통해 근로자를 찾는 최선의 결정을 예측할 수 있습니다. 지식 관리 및 데이터 마이닝은 정부 관리를 개선하는 주요 방법입니다. 빅데이터를 공공 관리와 통합하면 정부가 아동 복지를 관리하는 능력을 향상시킬 수 있습니다. 바이두의 CEO인 로빈 리 (Robin Li)는 2017년 NPC와 중국의 CPPCC에서 "실패 아동의 문제를 해결하기 위해 인공지능과 빅데이터 기술을 사용하는 제안"을 발표했습니다. 안면 인식 및 인공지능 기술의 빅데이터를 사용하여 어린이 유실 사례의 탐지 효율성 및 탐지 속

도를 크게 향상시킵니다. 그것은 잃어버린 아이들을 찾는 데 적합한 안면 인식 모델을 구축하고 전국을 커버하는 실종 아동에 대한 데이터베이스 구축 및 안면 인식 기술과 공공 보안 및 교통 모니터링 시스템을 결합하는 것을 포함합니다.

## 다. 빅데이터의 분석 기술 단계

행동을 예측하는 4가지 단계가 있습니다. 첫 번째는 행동 위험 지식을 확립하는 것입니다. 두 번째는 지능형 이미지 모니터링 및 모바일 애플리케이션을 통해 안전하지 않은 행동을 수집하는 것입니다. 세 번째는 의미 정보 이미지를 Hadoop 분산 파일 시스템을 통해 저장하는 것입니다. 네 번째는 행동을 분석하고 예측하는 것입니다.

빅데이터 분석 프로세스에는 6가지 단계가 있습니다. 첫째, 직접 및 간접 데이터에서 데이터를 수집하는 것입니다. 둘째, 통합 프로세스를 통해 빅데이터의 품질을 향상시키는 것입니다. 셋째, 통계적 방법과 기계 학습을 사용하여 학습 모델을 개선하는 것입니다. 넷째, 테스트 데이터로 모델을 평가하는 것입니다. 다섯째는 모델의 실제 적용이며 마지막으로 모델 성능 예측의 정확성을 감지하는 것입니다.

## 라. 정부 거버넌스의 빅데이터 기준

빅데이터의 기준에는 상황별 적절성, 운영 적절성 및 임시 적절성이 포함됩니다. 투명성, 유연성 및 책임성 기준을 사용하여 데이터 사용 및 사용자를 평가할 수 있습니다. 데이터 사용의 궁극적인 목표는 지표의 계산을 명확하게 정의하고, 빅데이터와 관련된 명백한 논쟁을 이해하고, 평가 기준을 명확히 하는 것입니다. 또한 민감한 정보의 보호가 점점 더 어려워지고 있기 때문에 정부는 통신 네트워크, 규제 표준 및 물류 디지털 시장의 표준을 개선하고 데이터 과학자와 같은 핵심 인재 부족을 채우기 위한 목표 프로그램을 개발해야 합니다. 개인 정보 보호와 기술의 결합을 통해 빅데이터는 개인 정보 보호 요구 사항을 충족합니다. 현재 정부는 사용자 개인 정보를 정의하기 위한 특별한 법률과 규정이 없으며, 관련 문제를 설명하기 위해 다른 관련 법률과 규정이 사용됩니다. 정부는 정책 수준에서 데이터 수집 사용자의 규

제를 명확하게 하고 개인 정보 보호의 정책 프레임 워크를 개선하고 개인 생활, 상업적 이익 및 사회적 거버넌스 사이의 가능한 균형을 추구하는 것이 시급합니다. 빅데이터 분석 과정에서 준수해야 할 원칙은 데이터를 확보하고 데이터를 분석하고 합법적이고 정기적으로 데이터를 적용하는 것입니다.

### 3. 정부 거버넌스에 대한 빅데이터의 기회, 문제점, 대책 및 연구 과제

#### 가. 정부 거버넌스에 대한 빅데이터 기회

빅데이터는 복잡하고 빠르게 증가하는 수많은 데이터 세트를 나타내며 많은 영역에 많은 기회를 제공합니다. 빅데이터 시대에는 데이터가 사회 과학 연구 개발에 도움이 되므로 데이터를 관리하는 것이 매우 중요합니다. 빅데이터는 정확한 예측을 가능하게 하고, 정보를 다른 출처와 통합하고, 지식을 창출하고, 더 나은 서비스를 제공하고 촉진합니다. 빅데이터는 중요한 정책과 정보를 전달합니다. 빅데이터와 조직의 효과적인 조합을 통하여 더 많은 정보를 얻을 수 있을 것입니다. 빅데이터를 처리함으로써 관리자는 비용을 줄이고 지속 가능한 개발을 달성할 수 있습니다. 빅데이터는 정부 문제의 투명성을 촉진하고, 정부의 신뢰성을 높이고, 정부 거버넌스 성과를 촉진하며, 정부 거버넌스의 수평 및 수직 격차를 좁혔습니다.

#### 나. 정부 거버넌스에 대한 빅데이터의 문제점

첫째, 개인 프라이버시 문제) 정부의 가장 큰 과제는 데이터 개인 정보 보호입니다. 데이터 수집, 결합, 분석시에 발생할 수 있는 문제입니다. 시민들은 개인 데이터를 제출했으며 추적 가능성을 모릅니다. 빅데이터를 보관하고 재사용하면 민감한 정보가 노출될 위험이 높아집니다. 각국 정부는 불법적인 개인 정보 거래에 대항하기 위한 노력을 반복하고 있으나 정부 부서 직원들에 의해서 개인 정보의 3/4 이상이 제공된다는 통계가 있습니다.

둘째, 빅데이터 사용이 부적절하다) 빅데이터의 세 가지 주요 과제는 "처리할 데이터, 분석 방법, 변경 방법"입니다. 맥킨지 연구소는 지난 2년간 정부가 90 % 이상의 데이터를 생성했지만 1 % 만 분석했다고 발표하였습니다. 또한 조직의 리더는 빅데이터의 가치를 통찰하는 능력이 부족합니다. 그리고



더 나은 결정을 내리기 위해 조직에 대한 업무 흐름과 인센티브 메커니즘이 없습니다. 데이터 분석을 위한 주변 환경이 척박하며 체계적인 관리가 부족합니다. 데이터 생성 프로세스는 연구 목적과 결합되지 않았으며, 데이터 품질은 신중하게 심사되지 않았으며 데이터 연결은 연구 공유의 목적을 달성하지 못했습니다.

셋째, 정부 데이터의 신뢰성) 정부가 수집한 데이터는 의심의 여지가 있습니다. 연구에 따르면 비즈니스와 사회 단체에서 일하는 18세에서 50세 사이의 여성의 데이터는 그들의 인식 부족으로 인해 신뢰도가 낮습니다. 반면에 노인들에게서 수집된 데이터는 신뢰할만한 가치가 있습니다. 그것은 고등 교육을 이수한 사람들은 데이터 수집에 대한 신뢰도에 민감하므로 더욱 신중을 기하기 때문입니다. 데이터 수집은 종종 다양한 기본 및 보조 데이터와 혼합됩니다. 또한 빅데이터의 등장에도 불구하고 일부 사회 학자들은 추측을 통하여 사실들을 호도하기도 합니다.

#### 다. 정부 거버넌스에 대한 빅데이터 대책

첫째, 개인 정보 보호 솔루션) 개인 데이터를 빅데이터와 분리하면 정책에 대한 데이터 분석이 더욱 강력 해집니다. 디지털 데이터 스트림을 분석함으로써 조직이 개발되지 않은 데이터로부터 이익을 얻을 수 있도록 지침이 제공되도록 합니다. 연구를 위한 커뮤니티가 국가간 데이터를 분석할 권리가 있는 체계를 만듭니다. 개인 정보의 남용을 보호하기 위해 조화로운 법적 체계를 확립하도록 합니다.

둘째, 데이터 사용 및 규제 문제) 데이터 사용의 효율성을 개선하기 위해 핵심은 인재를 발굴 및 유치하고, 적절한 인력과 기술을 올바른 곳에 배치하고, 워크 플로우와 인센티브 메커니즘을 변경하여 빅데이터의 사용을 극대화하고, 데이터 기술 분석을 개선하고, 조직의 적응성을 향상시키며 디지털 및 자동화를 포용하는 작업을 재설계합니다. 데이터 소스의 진위 여부를 판단함으로써 스코어 카드, 표준화 및 관리 규칙, 그리고 교육 직원을 활용하여 데이터 사용 효율성을 향상시킬 수 있습니다. 데이터를 활용하여 조직의 지속 가능한 개발을 강화하고 연구원들이 데이터를 탐색하고 연구할 수 있도록 합니다.

## 라. 연구 과제

첫째, 빅데이터를 완전히 활용하기 위해서는 정부가 빅데이터 법규를 더욱 개선 및 정제하고 정보 고립을 좁히고 정부 데이터를 공유해야 합니다. 데이터 수집을 개선할 수 있는 많은 인구, 모바일 인터넷 및 사물 인터넷이 있습니다. 빅데이터는 큰 학문적 가치, 응용 가치 및 개발 잠재력을 가지고 있으며, 빅데이터에 대한 법률과 규정의 신속한 개발이 요구됩니다.

둘째, 정부는 빅데이터 기술의 처리를 더욱 개선해야 합니다. 기존 기술은 중앙 집중식 스토리지 데이터에 중점을 두지만 실질적인 애플리케이션 가치는 거의 없습니다. 따라서 복잡한 구조 데이터, 반 구조화된 데이터, 비정형 데이터 연구가 강화되어야 합니다.

셋째, 정부는 데이터 플랫폼을 더욱 강화해야 합니다. 빅데이터는 정부 문제를 해결하는 좋은 방법이지만 기술 및 정책상의 이유로 개인, 사업 및 정부 개발에 심각한 영향을 미치거나 심지어 손상시킬 수도 있습니다. 정부는 데이터 수집, 분석, 저장 및 관리 플랫폼을 강화해야 합니다.

## IV. 미국 정부의 빅데이터 활용

### 1. 미국 연방 정부의 빅데이터

빅데이터는 그것의 초기부터 계속 주목을 받아 왔습니다. Web of Science 핵심 컬렉션(WoS)을 참조하면 빅데이터 쿼리는 2004년 컨퍼런스 논문에서 최초의 사용을 기점으로 2013년부터 비약적인 증가를 통하여 7천 개가 넘는 조회수가 발생하였습니다. 뚜렷한 지표로서는 특성별 빅데이터(2008), 의미별 빅데이터(2012), 정치 분석에 의한 정치 과학에서의 빅데이터(2013), 빅데이터 및 조직 디자인, 조직 디자인 저널(2014), 디지털 저널리즘에 의한 빅데이터 시대의 저널리즘 (2015), 정보 과학 저널에 의한 빅 소셜 데이터(2015), 비즈니스 및 경제 통계에 의한 농업 빅데이터의 가치(2016) ) 및 심리학, 심리학적 방법의 빅데이터(2016) 와 같은 다양한 분야에서 빅데이터 특별 이슈들이 게시되었습니다.

급속하게 증가하는 문헌의 양에도 불구하고, 빅데이터의 의미는 여전히 희미하게 나타납니다. 그정의들도 다양하지만 반드시 서로 일치하지는 않습니다. 더욱이, 그것은 다양한 정의의 표현들만이 아니라 특성들, 즉 Vs도 변합니다. 예를 들어, Gartner(2016)는 3V(즉, 볼륨, 다양성 및 속도)를 제안하였고, IBM은 3V에 유연성(2016)을, SAS는 가변성(2016)을, Oracle은 가치(2016)를 더하였습니다. 또한 빅데이터의 의미를 탐구하기 위한 연구들은 드문 것으로 보이며 소수의 기사(저널 및 컨퍼런스 논문 포함)의 초점은 빅데이터 분석/처리 또는 특정 영역에의 컨셉 적용에 관한 것입니다. 이러한 상황은 2016년 이후 빅데이터 개념에 대한 전담 연구를 통해 현재 연구의 분석이 완료된 후에도 계속되고 있습니다.

이러한 현실에서, 빅데이터에 대한 이해를 신속하고 적절하게 이해하는 방법은 컴퓨터 과학, 데이터 관리, 통계 등과 같은 빅데이터의 출현과 직접 접촉하지 않는 분야의 전문가에게는 여전히 어려운 과제가 되었습니다. 동시에 빅데이터가 궁극적으로 자신의 분야에 영향을 미칠 것이라는 점을 지속적으로 상기시키고 있습니다. 기존 빅데이터 정의를 조사하거나 다양한 수의 V를 합성하는 것은 이해를 얻는 방법 중 하나일 수 있을 것입니다.

## 가. 용어의 사용

빅데이터 라는 용어는 2010년 12월 미국 연방 정부에서 처음으로 대통령 과 의회에 보고서 “디지털 미래 : 네트워킹 및 정보 기술의 연방 자금 지원 연구 및 개발 보고서”라는 제목의 문서에 등장했습니다. 이 보고서는 과학, 기술 및 혁신 문제에 관한 정책 상담을 위해 백악관에 의해 임명 된 미국의 주요 과학자 및 엔지니어 자문 그룹인 과학 기술 자문위원회(President 's Council of Science and Technology)에서 작성했습니다. 이 보고서의 주요 목적은 1991년 미국 고성능 컴퓨팅법(High-Performance Computing Act), 1998년의 차세대 인터넷 연구법(the Next Generation Internet Research Act) 및 2007년의 미국 COMPETES 법(기술, 교육 및 과학의 우수성을 의미있게 홍보 할 수 있는 기회 창출)의 맥락에서 연방 정부가 자금을 지원하는 고급 정보 기술의 주요 원천인 연방 네트워킹 및 정보 기술 연구 및 개발(NITRD) 프로그램에 대한 평가를 제시 하는 것입니다. 2010년 평가를 위해 이 보고서는 NITRD 프로그램을 ‘연방 정부가 네트워킹 및 정보 기술 (NIT)에 대한 미분류 R&D 투자를 조정하는 메커니즘’ 그리고 ‘조정 노력이 아닌 연방 정부가 지원하는 NIT의 분류되지 않은 R&D 노력의 양상블’ 이라는 두 가지 측면에 중점을 두었습니다. 보고서에서 확인한 5가지 주요 주제 중에서 데이터 양(volume)이 기하 급수적 인 증가 때문에 첫 번째 항목으로 선정되었으며 모든 조직이나 기관에는 빅 데이터 전략이 권장되었습니다.

그 이후로 빅데이터라는 용어는 많은 중요한 연방 문서에서 처음에는 인용 부호와 모두 소문자로 표시되었지만 이후에는 인용 부호는 제거되고 첫 글자는 대문자로 표시되었습니다(예 : 2016년 연방 빅데이터 연구 개발 전략 계획). MGI는 1990년에 McKinsey & Company가 설립한 비즈니스 및 경제 분야의 전담 연구 기관인 McKinsey Global Institute의 약자입니다. 맥킨지 글로벌 연구소는 미국 연방 기관이 아니지만 2011년 보고서 ‘빅 데이터 : 혁신, 경쟁의 다음 경계, 그리고 생산성’은 보고서의 첫 번째 저자인 James Manyika가 2012년 3월 29일 워싱턴 DC의 한 강당에서 열린 연방 정부 빅데이터 발표회에 참석했기 때문에 포함되었습니다. 발표회는 빅데이터 프로젝트의 대행사 발표와 패널 토론으로 구성되었으며 Manyika

는 패널리스트 중 한 명이었습니다. 발표회는 2012년 미국 연방 정부 빅데이터 이니셔티브에 대한 과학 기술 정책실 발표에 즉각적인 후속 조치였으므로 MGI 보고서의 영향을 간과해서는 안 됩니다. 과학 기술 정책실은 과학 기술 자문위원회를 관리하는 백악관 행정실의 한 부서입니다. 빅데이터 이니셔티브의 과학 기술 정책실 발표와 빅데이터 팩트 소책자가 본 연구의 초점이었습니다. 빅데이터 IWG는 2011년 초에 형성된 빅데이터 SSG(Senior Steering Group)로 알려진 빅데이터 기관간 중재를 위한 실무 그룹을 지칭하며 연방 정부의 빅데이터 연구 활동을 식별하는 작업을 수행합니다. 2015년 국립 표준 기술 연구소 빅데이터 공공 실무 그룹(Public Working Group)은 빅데이터를 “전통적인 데이터 아키텍처가 새로운 데이터 셋을 효율적으로 처리 할 수 없음” 으로 정의하는 7권 시리즈 “빅데이터 상호 운용성 프레임 워크”의 첫 번째 볼륨을 발표했습니다. 이미 위에서 이야기한 것과 마찬가지로 빅데이터를 정의하는 이 방법은 기본적으로 전통적인 데이터 아키텍처에 대한 지식이 없는 전문가에게는 도움이 되지 않았습니다.

## 나. 빅데이터 이니셔티브

2010년 과학 기술 자문위원회 보고서의 권고에 따라 미국 연방 정부는 2012년 3월 29일에 빅데이터 연구 개발 이니셔티브를 공식적으로 발표하여 빅데이터 R&D에 최대 2억 달러를 투자하여 빠르게 증가하는 디지털 데이터와 국가의 가장 시급한 과제 해결에 도움을 주기로 결정하였습니다. 이 발표에서 국가 과학 재단, 국립 보건원, 국방부, 에너지부, 국방 고급 연구 프로젝트 기관 및 내무부의 미국 지질 조사국을 포함한 6개의 연방 부서와 기관이 주요 대표자로 발표하였습니다. 6개의 부서와 에이전시는 같은 날 오후에 각 웹사이트에서 결정한 사항에 관한 보다 자세한 정보를 제공했습니다. 예를 들어 국가 과학 재단은 웹 사이트에 16개의 빅데이터 프로그램을 일부 프로젝트에 대한 설명과 다른 프로젝트에 대한 핫링크와 함께 설명했습니다. 빅데이터 이니셔티브를 발표했다고 해서 연방 정부에서 빅데이터 조치가 실제로 시작된 것은 아니며 빅데이터라는 용어를 공식적으로 사용하는 것이었음을 숙지해야 합니다. 발표 당시 이미 진행중인 많은 빅데이터 프로그램과 프로젝트가 있었으며 빅데이터 팩트 소책자에 명확하게 표시되어 있었

습니다. 발표일과 동일한 날에 과학 기술 정책실이 발표한 팩트 소책자에는 1997년에 시작된 NASA의 지구 과학 데이터 및 정보 시스템과 2001년에 시작된 에너지부의 차세대 네트워킹 프로그램을 포함하여 연방 정부에서 진행 중인 총 89개의 빅데이터 프로그램 및 프로젝트를 통해 확인되었습니다. 과학 기술 정책실 2012 발표 또는 과학 기술 자문위원회 2010 보고서에서는 빅데이터에 대한 정의를 제공하지 않았지만 ‘대형 및 복합 디지털 수집’과 ‘기하 급수적으로 증가하는 데이터 및 데이터 볼륨’과 같은 설명 정보를 제공했습니다. 맥킨지 글로벌 연구소 2011 빅데이터 보고서 및 국가 과학 재단의 빅데이터 홍보 자료와 같은 비교적 덜 권위적인 기관의 문서에서 정의가 발견되었습니다. 맥킨지 글로벌 연구소는 빅데이터를 ‘크기가 일반적인 데이터베이스 소프트웨어 도구로 캡처, 저장, 관리 및 분석 할 수 있는 능력을 넘어서는 데이터 세트’로 간주했으며 국가 과학 재단 빅데이터 홍보 자료에서는 빅데이터를 ‘현재 그리고 미래에 이용 가능한 도구, 센서, 인터넷 거래, 이메일, 비디오, 클릭 스트림 및 기타 모든 디지털 소스에서 생성되는 대형, 다양, 복합, 중단 또는 분산 데이터 세트’로 정의하였습니다. 또한 미국 연방 정부가 처한 환경과도 관련이 있다고 간주되는 또 다른 정의는 TechAmerica 재단의 연방 빅데이터 위원회에서 생성한 보고서에서 발견되었습니다. 보고서는 ‘빅데이터는 대량의 고속, 복합 및 가변 데이터를 설명하는 용어로서 정보의 캡처, 저장, 배포, 관리 및 분석을 가능하게 하는 고급 기술과 기술이 필요합니다’라고 설명하고 있습니다.. 외부 상황과 마찬가지로, 이러한 정의는 유사성이 존재하더라도 그 표현이나 의도에서 서로 일치하지 않습니다. 2014년 빅데이터 검토 보고서에 언급된 바와 같이, 빅데이터 정의는 풍부하며 이를 정의한 당사자에 따라 달라집니다. 과학 기술 정책실의 소책자에 등재된 89개의 빅데이터 프로그램 및 프로젝트는 연방 정부의 빅데이터 조치의 강점으로 간주되고 있습니다.

## 2. 빅데이터와 정부의 변화

### 가. 빅데이터 활용

미국 연방 정부는 3개 부문(집행, 입법 및 사법)으로 구성되어 있으며, 이

모두는 많은 양의 데이터를 생성합니다. 여기서는 경영진의 데이터에 중점을 둡니다. 표는 15개의 연방 행정 부서와 이러한 부서에 대한 빅데이터를 생성하는 일부 소스가 나와 있습니다. 예를 들어, 재향 군인국의 직원은 312,878 명입니다. 또한 전자 정부 플랫폼을 통해 재향 군인국에 액세스하면 매달 2,700,000 명의 사용자가 재향 군인국 웹사이트를 방문하여 대량의 데이터를 생성합니다. 페이지에 소요된 시간, 방문 빈도, 방문당 완료된 트랜잭션 등과 같은 데이터를 수집하면 웹 분석의 필요성이 높아질 것입니다.

연방 정부 행정부의 빅데이터 소스

부서	웹사이트	직원수	월 사이트 방문수	사이트 링크수
농무부	usda.gov	98,235	2,800,000	69,658
상무부	commerce.gov	45,348	7,700	42,281
국방부	defense.gov	771,614	441,100	10,237
교육부	ed.gov	4,611	2,900,000	53,467
에너지부	energy.gov	16,651	501,400	24,334
보건복지부	hhs.gov	83,745	173,200	35,950
주택도시개발부	hud.gov	9,818	789,700	32,376
국토안보부	dhs.gov	191,197	680,600	21,280
내무부	doi.gov	72,168	41,100	6,550
법무부	justice.gov	118,104	446,000	19,488
노동부	dol.gov	16,554	547,400	23,760
국무부	state.gov	12,086	1,800,000	61,586
교통부	dot.gov	58,189	841,300	36,432
재무부	treasury.gov	112,541	308,100	8,053
보훈부	va.gov	312,878	2,700,000	33,798

## 1) 웹 분석 활용

웹 분석은 온라인 사용자와 조직의 상호 작용에 대한 포괄적 개요를 제공합니다. 비즈니스 웹 사이트에서는 2% 미만의 방문자가 구매하지만, 모든 방문자는 정보를 제공합니다. 이러한 사이트는 방문자의 방문 지점, 페이지 뷰, 페이지 방문 시간, 버려진 장바구니, 다운로드 및 검색수와 같은 데이터를 수집합니다. Alexa, Compete, Comscore, Google 및 Quantcast와 같은 웹 기반 회사는 사이트 트래픽, 글로벌 및 로컬 순위, 사이트 평판 등 웹 분석에 유용한 데이터를 포함한 다양한 웹 메트릭을 제공합니다. 서비스를 제공하는 정부 웹 사이트에 대해서도 유사한 유형의 정보를 수집해야 합니다. 이러한 웹 사이트는 일반적으로 상용 사이트보다 트래픽이 훨씬 적지만 웹 분석을 사용하여 수집된 데이터에서 충분한 정보를 추출하여 고객 서비스를 향상시킬 수 있습니다. 어떤 정보가 소비되고 있는지 분석함으로써 정부 플랫폼의 효과를 모니터링하기 위한 지표를 개발할 수 있습니다.

## 2) 정부 변화의 추진

정부 기관은 전자 정부를 통해 공공 부문 서비스를 개선하기 위해 정보 통신 기술을 채택하고 있습니다. 정부의 변화는 제품 혁신, 고객 서비스 제공 개선, 공공 부문 인프라 관리 개선 및 운영 비용 절감으로 이어질 수 있습니다. 정부의 변화는 다음을 통해서 이루어질 수 있습니다.

- 집계 : 정보 흐름을 촉진하기 위해 다른 정부 기관과 협력 제휴를 맺습니다.
- 신디케이션 : 핵심 역량을 활용하여 시민 또는 다른 기관에 대한 서비스 제공을 향상시킵니다.
- 소비 : 외부 데이터 소스로 대학 및 비즈니스와 같은 다른 엔티티를 참여시킵니다.
- 공동 창조 : 외부 파트너와 혁신을 통해 프로세스 및 지적 재산 관리를 개선시킵니다.

정부가 변화하면 많은 이점이 있지만, 디지털 분배로 인한 액세스 제한, 속



련된 인력의 가용성 제한, 높은 비용 및 레거시 시스템과의 통합 문제 등 기술 구현 및 사용과 관련된 많은 문제가 발생합니다. 이러한 모든 것은 또한 조직적, 정치적 이슈로 인해 변화하기 쉽습니다. 한 가지 방법은 클라우드 아키텍처와 같은 대체 플랫폼으로 데이터를 마이그레이션하고 빅데이터 세트를 분석하여 의사 결정을 개선하는 것입니다.

## 나. 빅데이터 소비

정부 웹 사이트를 전자 상거래 웹 사이트와 비교할 때 전자 정부 사이트가 뒤쳐져있는 경우 성능 차이가 있습니다. 정부가 빅데이터에서 가치를 끌어내려고 노력함에 따라, 빅데이터에서 가치를 도출하는 측면에서 민간 부문에 비해 뒤쳐질 위험이 항상 존재합니다. 이러한 기술 발전의 지연은 또 다른 중요한 문제, 즉 정부 부문의 데이터의 동시 과잉 생산 및 과소비 문제를 야기합니다. 이것은 그러한 상황과 관련된 잠재적인 가치와 본질적인 도전을 어떻게 설명 할 수 있는지에 대한 의문들을 불러 일으킵니다.

### 1) 과잉 생산과 부족

과잉 생산과 과소비에 대한 경제 이론은 연방 정부를 개선하기 위해 빅데이터 분석을 사용하는 것의 잠재력과 도전을 모두 조율하는 데 도움이 됩니다. 과잉 생산이란 때때로 생산의 분배가 불안정하고 생산 활동의 장소간 분배의 불평등을 의미합니다. 물리적 재화의 지속적인 과소비와 함께 이 과잉 생산은 일반적으로 가격 하락과 제품 평가 절하를 특징으로 하는 위기 상황을 초래할 것입니다. 그러나 빅데이터는 이러한 경제적 근거에 이상을 나타냅니다. 이 경우 소비 수준이 낮기 때문에 소비되지 않은 데이터가 더 많이 발생합니다. 더 많은 데이터가 생성될수록 데이터 손실 가치 대신 이러한 데이터를 분석하여 얻을 수 있는 잠재적 통찰력의 가치가 높아집니다. 적절한 분석을 통해 빅데이터를 유용한 지식과 수량화 가능한 수치들로 변환함으로써 데이터의 과잉 생산으로 인한 이 잠재적 이익에 액세스할 수 있으며 데이터 생산 및 소비 비율이 더 유사한 비율로 나타납니다. 빅데이터 소비가 생산 속도보다 계속 지연될 수 있지만 분석 기회로 인해 가치가 계속 증가할 수 있으며 정보가 기술적으로 처리되고 조직적으로 관리되는 방식에 영향을 미칩니다.

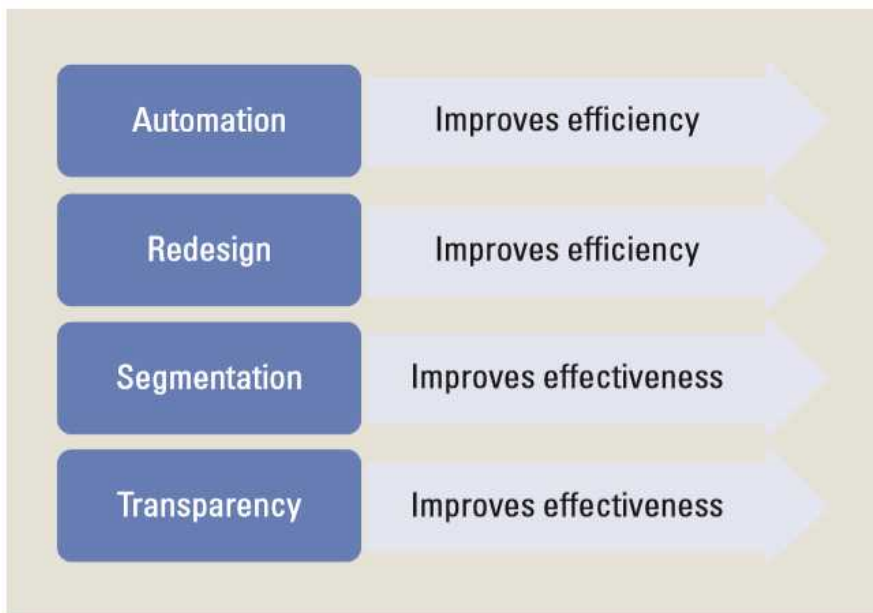
## 2) 미국 재향 군인국

예를 들어, 미국 재향 군인국을 고려해 보겠습니다. 이라크와 아프가니스탄에서의 전쟁이 중단되면서 다른 재향 군인들이 서비스 수준에 따라 혜택을 처리하는 데 시간이 오래 걸리는 경우가 종종 있습니다. 재향 군인국의 보유중인 클레임의 수집은 2009년 400,000개에서 2012년 880,000개로 증가했으며, 125일이 지난 장애 클레임은 2010년의 180,000개에서 2012년 말까지 594,000개로 증가했습니다. 재향 군인국의 처리 장치에서, 재향 군인이 재향 군인국에 데이터를 계속 제출함에 따라 빅데이터 분석을 통해 명백한 처리 백로그를 줄일 수 있는 기회가 있습니다. 예측 및 규범적 분석을 사용하여 각각의 새로운 클레임 처리를 개선하기 위해 새로운 기준들을 생성하여 규모의 경제를 창출할 수 있습니다. 재향 군인국에서 데이터 관리시에 직면한 또 다른 당면 과제는 종이 기반 처리의 확산입니다. 이러한 잉여 데이터는 자동화 및 일부 내부 기능의 재설계로부터 즉시 혜택을 받을 수 있습니다. 빅데이터 분석은 재향 군인국의 클레임 처리 기술에 급격한 변화를 일으킬 수 있습니다. 더 큰 데이터가 분석됨에 따라 퇴역 군인의 여러 세그먼트간에 상관 관계를 명확하게 보여줄 수 있는 패턴이 발생합니다. 빅데이터 분석은 또한 미국 국방부와 재향 군인국간의 협력을 강화할 수 있습니다. 예를 들어, 재향 군인국의 목표는 주로 국방부가 현역, 방위군 및 예비군을 마치는 모든 서비스 및 직원 레코드에 전자 액세스를 제공하도록 함으로써 재향 군인국에 도착하는 완전히 처리된 클레임 수를 3%에서 20%로 늘리는 것입니다. 완전히 처리된 클레임에는 서비스 기록 및 출입국 시험과 같은 국방부의 모든 관련 정보가 포함됩니다.

### 다. 효율적이고 효과적인 서비스

정부의 효율성이 높아짐에 따라 과제를 완료하는 데 필요한 시간과 노력이 줄어듭니다. 정부 외부에서는 데이터 중심 기업이 경쟁 업체보다 우수한 성능을 보이며 평균 5% 생산성과 6% 수익성을 향상시켰습니다. 예를 들어 소매업 체인 Sears는 빅데이터 분석을 사용하여 광고 프로모션 개발 시간을 8주에서 1주 이하로 줄였습니다. 관련 내부 조직이 빅데이터를 수집하고 적절하게 분석할 수 있으면 보고서를 작성하는 데 걸리는 시간이 줄어들고 보

다 구체적인 종류의 추가 분석이 실행됩니다. 또한 빅데이터 소비 및 분석을 통해 작업을 처리하는 데 필요한 노력도 줄여야 합니다. 빅데이터는 전자 정부 영역에서 성과를 측정하기 위한 특정 지표를 제공할 수도 있습니다. 우리는 데이터 분석 프로세스를 자동화하고 재 설계함으로써 정부 효율성을 높이기 위해 빅데이터를 사용해야 합니다. 또한 빅데이터 분석은 데이터 세분화 및 정보 투명성을 통해 정부의 효율성을 높일 수가 있습니다.



빅데이터를 활용하여 정부를 변화시키는 모델.

### 1) 오토메이션

자동화는 비즈니스 및 정부에서 정보 통신 기술을 구현하는 초석 중 하나입니다. 병목 현상을 대상으로 빅데이터를 간소화하고 분석을 지원할 수 있습니다. 예를 들어, 재향 군인국 사례의 경우, 서류 제출 사례의 백로그는 데이터 자동화 기회를 나타냅니다. 종이 제출을 전자 기반 제출로 옮기면 일부 데이터가 자동으로 처리되므로 청구를 처리하는 데 필요한 시간이 줄어듭니다.

### 2) 재설계

유전자 알고리즘, 회귀 분석 및 감성 분석 도구와 같은 빅데이터 분석을 사용하여 프로세스를 재설계 할 수 있습니다. 회귀 분석은 잘 정의된 변수 간

의 관계를 조사하고, 감성 분석은 의견 진술에서 긍정/부정 등의 성향을 추출하려고 합니다. 이러한 유형의 분석 도구는 직원들이 고객의 요구를 더 잘 이해하도록 함으로써 서비스 제공을 향상시킬 수 있습니다. 예를 들어, 미국 국세청은 빅데이터 분석을 통합하여 사기 탐지를 개선하고 부적합 사항을 발견하도록 세금 신고 프로세스를 재설계했습니다. 빅데이터 분석과 그로부터 파생된 정보의 신중한 사용이 다른 연방 부서와 다른 정부 기관에서 증가함에 따라 거래 처리에 더 적은 시간과 노력이 필요하게 되었습니다.

### 3) 분할

세분화는 특정 데이터 클러스터 또는 그룹을 나타냅니다. 이것은 일반적인 마케팅 개념으로, 인구 통계 또는 지역을 기반으로 그룹을 만드는 데 자주 사용됩니다. 공공 부문에서 빅데이터를 세분화하면 간단한 데이터 검사만으로 직관적이지 않거나 쉽게 볼 수 없는 클러스터가 드러날 수 있습니다. 빅데이터 분석을 통해 공무원은 다양한 관점에서 데이터를 읽고 새로운 정보를 공개 할 수 있습니다.

재향 군인국에 제출된 클레임에서 집계된 빅데이터의 사례를 고려하여 특정 의료 상태와 같은 클러스터 전체에 숨겨진 링크를 보여줌으로써 재향 군인국이 신속하게 클레임을 처리 할 수 있는 새로운 표준을 만들 수 있습니다.

### 4) 투명도

빅데이터 도구는 대량의 데이터에 대한 보고를 쉽게 지원할 수 있으므로 더 많은 정보를 대중에게 제공할 수 있습니다. 예를 들어, 자치 정부 영역에서 소셜 미디어의 성장은 이미 정부의 일부 영역에서 투명성을 높이고 부패를 줄였습니다. 정부 변화에 따라 빅데이터 분석이 증가함에 따라 행정의 효율성을 증가시키면서 의사 결정은 추측에 의하기 보다는 데이터에 의해 발생합니다. 개방된 정부라는 개념은 더 많은 정보를 대중에게 공개해야 합니다. 개방성에 대한 기대치가 높아짐에 따라 의사 결정자는 데이터 입력을 기반으로 결과를 정당화해야 합니다.

## 라. 빅데이터 장벽 극복

정부가 빅 데이터 분석을 사용함으로써 정부의 변화가 도출될 수 있지만, 여전히 몇 가지 과제를 해결해야 합니다.

#### 1) 비정형 데이터 분석

첫째, 기존의 관계형 데이터베이스 시스템은 이미지, 블로그, 스마트 폰, GPS, 모바일 장치 및 소셜 네트워크와 같은 비정형 소스에서 수집한 빅데이터를 관리하는 데 적합하지 않습니다. 오늘날 생성되는 데이터의 약 85%는 비정형입니다. 정부는 비정형 데이터의 조작 및 분석을 지원할 수 있는 데이터 저장소를 유지해야 합니다.

#### 2) 인프라 구축

둘째, 정부 부서와 기관이 빅데이터를 관리하기 위해 적절한 기술 인프라를 채택해야 합니다. NASA 및 미국 국세청과 같은 기관은 웨어 하우스 최적화, 스트리밍 데이터, Hadoop 및 기타 기술을 사용하여 빅데이터를 관리합니다. Hadoop은 대규모 컴퓨터 클러스터를 사용하여 소매에서 생물 정보학에 이르는 산업의 빅데이터를 분석하는 오픈 소스 솔루션입니다. 실시간 요청으로 인해 데이터 소비 요구가 증가함에 따라 기존 플랫폼이 해결하지 못하는 격차를 해소 할 수 있는 오픈 소스 및 독점 제품이 추가로 등장하고 있습니다.

#### 3) 변경 수락

정부 부서는 전통적으로 가파르고 변화의 바람에 저항 할 수 있는 대규모 관료제입니다. 조직 관성은 새로운 아이디어와 새로운 방법론의 성장을 방해할 수 있습니다. 그러나 데이터의 가치가 공개됨에 따라 더 많은 옹호자들이 빅데이터의 가치를 공공 행정 개선 도구로 볼 것입니다. 공공 부문 기관에서 선거주기는 변화 속도와 새로운 운영 방식 채택에 영향을 미쳐 현재 기술의 채택 속도를 높이거나 구현에 필요한 지출을 늦출 수 있습니다.

#### 4) 개인 정보 문제 해결

마지막으로 개인 정보 보호 문제로 인해 공용 도메인에서 빅데이터를 채택하지 못할 수 있습니다. 민감한 데이터를 수집하고 조작하는 것은 정부 내

부 및 외부의 많은 그룹이 관심을 갖는 주제입니다. 정부 부서는 다른 정부 기관과 독점으로 간주하는 데이터를 공유하고 싶지 않을 수 있습니다. 이 잠재적 장벽을 줄이기 위해서는 감독과 적절한 관리가 필요합니다.

### 3. 미국 주정부의 무결성 평가 사례 연구

완벽한 사회적 청렴 시스템에는 정부 무결성, 기업 무결성 및 개인 무결성이 포함됩니다. 정부 무결성이 핵심이며, 기업 무결성이 중심이며, 개인 무결성이 기초입니다. 정부 무결성은 정부의 정당성, 좋은 지배 구조의 초석, 정부의 신뢰의 근원 및 토대를 전제로 합니다. 행정 기관과 직원의 청렴한 행동은 다른 사회 구성원의 의견, 복종 및 경쟁을 유발하며 전체 사회 분위기 형성에 큰 영향을 미칩니다. 정부 무결성의 핵심 가치에는 약속 이행, 공정성 및 정의, 투명성 및 개방성, 책임 및 효율성 등이 포함됩니다. 이에 따라 정책의 빈번한 변경, 약속 이행 불이행, 나태, 법준수 실패, 약한 법집행, 부채 불이행, 재량 남용, 공무원 부패 등과 같은 다양한 형태의 정부 청렴 범죄가 있습니다. 그래서, 정부가 무결성 거버넌스를 적극적으로 변화시키고 실행하도록 장려하기 위해 정부 무결성에 대한 객관적이고 공정한 이해와 과학적 평가를 수행할 필요가 있다. 그러나 다른 정부 기관들 사이에서 청렴의 의미가 완전히 일치하지는 않습니다. 그리고 사람들의 의견 사이에는 격차가 있습니다. 따라서 신뢰할만한 평가를 얻는 것은 매우 어렵습니다. 빅데이터 시대에는 빅데이터를 기반으로 정부 무결성 평가를 수행하는 것이 불가피한 추세입니다.

주정부의 무결성은 주로 세 가지 측면에서 조사되었습니다. 첫 번째는 존재 여부인데, 이는 공공 책임을 장려하고 부패를 억제하기 위한 법률 및 시스템을 포함한 공공 무결성 메커니즘의 존재를 조사하는 데 사용됩니다. 두 번째는 이러한 메커니즘의 효과를 확인하는 것입니다. 세 번째는 시민들이 이러한 메커니즘에 접근할 수 있는지의 여부입니다. 각주별로 정부 책임과 투명성과 관련된 14개 범주와 61개의 하위 범주에서 330건의 무결성 지표를 생성하여 점수화 하였습니다. 점수화하는 주요 목적은 무결성 프레임 워크의 부적합성을 인식하고 정부가 직면한 문제를 파악하며 정부가 개혁을 수행하도록 장려하는 것입니다.

## 주 정부별 무결성 평가 점수



## 가. 정부 무결성 평가의 이해

정부 무결성 평가는 부패에 대한 대중의 인식과 관심을 높이고 무결성이 결여되어 있음을 인식하고 적극적으로 개혁을 수행하도록 정부를 독려하는 데 매우 중요한 역할을 합니다. 이러한 관행은 더욱 다양하고 복잡해 지지만 일반적인 단점을 공유합니다. 대부분의 무결성 평가는 주로 전문가, 언론인, 사업가, 공무원, 공공 서비스 사용자 또는 무작위로 선택된 시민의 설문 조사에서 얻은 지각적인 데이터를 기반으로 합니다.

### 가) 데이터 수집 방법

설문 조사를 통한 데이터 수집 방법은 간단하고 구현하기 쉽지만 고유한 결함이 있습니다.

- 강한 주관성: 부패에 대한 인식과 부패 자체는 같은 것이 아닙니다. 설문 조사에 의해 수집된 데이터는 부정확하거나 오해의 소지가 있거나 자기중심적이며 메아리방 효과(echo chamber effect)에 의해 쉽게 영향을 받을 수 있습니다. 사람들은 같은 사건에 대해 다른 의견을 가질 수 있습니다. 어떤 사람들은 고위 공무원의 유죄 판결이 부패의 큰 증가를 의미한다고 주장하는 반면, 다른 사람들은 이것이 개혁이 마침내 시작되었다는 증거라고 확신할 수 있습니다.
- 불충분한 내용: 설문 조사에 의해 수집된 데이터는 부패가 가장 심각한 곳을 지적할 수 있지만 부패를 다루는 방법에 대해 정책 입안자들에게 제공한 지침은 제한적입니다.
- 설문 내용과 사용자의 피드백에 대해 균일한 표준을 설정하기가 어려워 샘플 편차가 발생하기 쉽습니다.

따라서 정부 무결성 거버넌스를 개선한다는 관점에서 무결성 설문 조사는 가치가 낮습니다. 정부 무결성 평가에는 보다 상세하고 객관적이고 양적이며 이해할 수 있는 데이터가 필요합니다.

### 나) 단일 인덱스 문제

국가 차원의 무결성 지수는 여러 정부 기관에서 발생하는 다양한 유형의



무결성 평가를 수치로 압축합니다. 경우에 따라 이러한 평가의 가치에 의문이 제기될 수 있습니다. 대부분의 평가가 너무 모호하여 실제 무결성 문제와는 거리가 멀 수 있습니다. 예를 들어, 일부 국가는 높은 수준의 부패를, 다른 국가는 정치 또는 관료적 계층 구조의 하위 수준에서 부패를 발견하고, 다른 국가는 선거 정치 및 후원에서 가장 심각한 부패를 발견합니다. 때때로 부패는 국제 무역 또는 초 국가적 범죄에 초점을 맞출 수도 있고, 다른 경우에 있어서는 국내 부패에 맞출 수도 있습니다. 따라서 단일 평가 결과는 정부 무결성 평가의 유일한 목적이 아니며 다음을 포함해야 합니다.

- 주관적 조사 데이터가 아닌 많은 객관적이고 정량적인 데이터를 수집하는 것이 정부의 의사 결정과 개혁을 더 잘 지원할 수 있습니다.
- 객관적이고 실제적인 데이터를 분석하고 정부의 특정 작업을 안내하기 위해 합리적이고 명확하며 설명 가능한 평가 표준 시스템을 구축합니다.

## 나. 빅데이터를 기반으로 한 정부의 무결성 평가

### 1) 목적 및 의의

#### 가) 정부 무결성 평가의 신뢰성 증진

적절한 거버넌스는 올바른 평가가 필요합니다.. 정부 무결성 평가는 정부 무결성 관리 개선을 촉진하는 중요한 메커니즘입니다. 정부 내부와 외부의 다양한 데이터 소스를 결합함으로써 폭과 깊이 있는 평가 지표가 확대됩니다. 그리고 적절한 평가 방법을 채택함으로써 평가 결과의 과학적 특성, 정확성 및 효과가 향상될 것입니다.

나) 정부 무결성 평가는 일반 평가에서 세부 사항에 대한 정확한 통찰력으로 변했습니다.

즉흥적인 점수 대신 정부 공무원의 객관적이고 정확한 행동 데이터 분석을 통해 관리자는 실제 변화를 발견하고 거버넌스의 복잡성을 포착할 수 있습니다.

니다. 무결성의 모든 측면은 이해하기 쉬운 측정값으로 평가할 수 있어 의사 결정자는 무결성 위협의 규모와 위치를 합리적으로 추정할 수 있습니다. 시간이 지남에 따라 무결성 위협의 경향을 상세하게 추적할 수 있으며, 이는 기존의 무결성 평가방법에서는 수행하기 어려웠던 것입니다.

#### 다) 정부의 무결성 거버넌스 능력 향상

의사 결정자가 무결성을 평가하고 청렴의 위험을 식별하는 데 도움을 줄 수 있습니다. 가장 중요한 것은 객관적이고 정확한 데이터 분석은 의사 결정자가 명백한 문제가 있는 섹션을 식별하는 데 도움이 될 뿐만 아니라 취해야 할 단계를 직접 지적하는 데 도움이 될 수 있습니다.

## 2) 빅데이터를 기반으로 한 정부 무결성 평가의 기본 원칙

빅데이터를 기반으로 한 정부 무결성 평가는 다음 기본 원칙을 따라야 합니다.

- 상세하고 포괄적이며 정확한 데이터여야 합니다.
- 평가는 즉흥적인 의견이 아닌 검증 가능한 증거를 기반으로 해야 합니다.
- 데이터 수집 관행이 정부 기관의 업무에 영향을 미치지 않아야 합니다.
- 정책적으로 중립이 유지해야 합니다. 그렇지 않으면 진행하기가 어려울 것입니다.
- 반복적인 평가가 필수적이며 측정 비용이 저렴해야 합니다.
- 선택한 데이터가 유효하고 검증 가능해야 하며 선택한 방법은 평가 결과의 신뢰성을 보장해야 합니다.
- 공무원, 시민 및 시민 사회 그룹은 무결성 평가를 쉽고 정확하게 해석할 수 있어야 합니다.
- 즉흥적인 설문 조사를 기반으로 한 기존 무결성 평가는 변경에 대한 유효한 정보를 거의 또는 전혀 제공하지 않습니다. 그러나 거버넌스는 시간이 지남에 따라 정책을 조정하는 효과를 추적하고 평가해야 합니다.
- 정부 내에서 또는 정부와 시민 사이의 불신을 조장하거나 공무원이 누군가가 끊임없이 자신의 어깨 너머로 보고 있다는 느낌을 주는 환경을 조성하는 경우 무결성 조치는 성공할 수 없습니다.

- 평가는 상황이 나쁘거나 좋은지, 개선 또는 악화되고 있는지를 알려줄 뿐만 아니라 성공할 가능성이 있는 개선 및 개혁을 직접 지적해야 합니다.

### 3) 빅데이터를 기반으로 한 정부 무결성 평가 작업 과정

정부 무결성 평가는 매우 민감하고 지속적이며 복잡한 시스템 엔지니어링입니다.

#### 가) 평가 주제

관련된 내부 정부 데이터로 인해 정부 무결성 평가는 정부 무결성 관리 기관이 주도할 가능성이 높습니다. 그리고 제 3의 기관은 데이터 분석, 표준화된 정보 시스템 구축 등과 같은 기술 서비스를 제공합니다.

#### 나) 평가 대상

평가 대상은 모든 수준의 정부 또는 정부 기관을 말합니다.

#### 다) 평가 지표 시스템 및 지표 데이터

데이터 지원이 없는 평가 지표 시스템은 의미가 없습니다. 따라서 이 시스템은 데이터 상황에 따라 공식화 되어야 합니다. 그러나 우수한 평가 모델은 가능한 한 포괄적이고 공정해야 합니다. 모델이 특정 분야에만 초점을 맞추면 모델 측에서 편향이 발생하기 때문입니다. 따라서 보다 포괄적인 데이터 소스를 사용해야 합니다. 정부 무결성과 관련된 데이터 소스는 주로 정부 활동 행동 데이터, 불만보고 데이터 및 인터넷의 여론 데이터를 포함합니다. 특히 인터넷의 여론 데이터를 사용하여 기존의 설문 조사 데이터를 대체할 수 있습니다. 왜냐하면 인터넷에 대한 여론은 특별히 감시된 데이터가 아니라 보다 현실적이고 객관적이며 샘플 크기 또한 크기 때문입니다.

#### 라) 평가 방법

평가 방법의 선택은 무결성 평가에서 중요한 문제입니다. 실제로, 평가 방법은 다양한 데이터 상황 및 평가 시나리오에 따라 조정되어야 합니다. 영구적인 최적의 방법은 없지만 특정 상황에 대한 최적의 방법은 있습니다.

다) 평가의 적용

적용은 정부 무결성 평가의 초석입니다.

바) 평가 추적 및 최적화

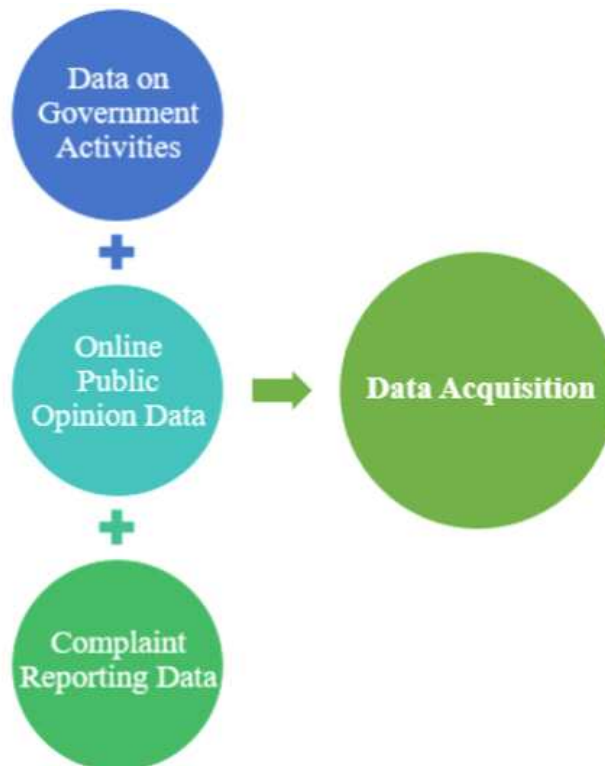
평가 시스템 자체도 최적화해야 합니다. 그래서 무결성 평가를 추적하고 평가해야 합니다.

사) 지속적인 평가

정기적인 평가 시스템을 구축해야 합니다. 1년에 한 번 정부 무결성 평가를 수행할 것을 권장하며, 이는 연말 보고서 및 성과 평가와 결합될 수 있습니다.

#### 4) 정부 무결성 데이터 수집 및 평가 지표

정부 무결성 평가의 핵심은 효과적이고 신뢰할 수 있으며 정확한 데이터를 수집하고 평가 지표를 결정하는 것입니다. 정부의 신뢰도를 강화하는 관점에서 데이터의 유효성 또한 강조됩니다.



정부 무결성 평가시 수집 데이터

## 가) 정부 활동에 대한 데이터

전자 정부의 급속한 증가는 정부 활동에 대한 객관적인 데이터 수집을 위한 중요한 기회를 제공한다. 현 정부의 비즈니스 시스템에서 공무원의 의사 결정 및 작업 기록에 대한 객관적이고 정량적인 데이터를 바탕으로 수학적 통계를 통해 효과적이고 신뢰할 수 있는 평가 지표의 계수 또는 측정을 수행할 수 있습니다. 그러나 중요한 질문은 어떤 지표를 수집해야 하는가입니다. 수집된 특정 지표 데이터는 현재 데이터 상황, 획득 어려움, 교체 가능 여부 및 기타 요인을 고려해야 합니다.

### ① 데이터 수집에 대한 생각

정부 무결성은 공익을 촉진하기 위해 정의와 형평성의 핵심 가치를 구현합니다. 투명성과 개방성 책임 그리고 효율성 등이 정부의 무결성과 관련된 데이터가 될 것입니다. 현재, 많은 연결되지 않은 사일로처럼 많은 정부 시스템 사이에 데이터 격차와 정보의 섬이 존재합니다. 따라서 가장 경제적인 데이터 수집 방법은 데이터와 인터페이스 표준을 통일하고 각 사일로시스템 내에 정규화된 통계 분석 모듈을 구축하며 데이터 인터페이스를 통해 데이터 중앙화 부서로 데이터를 정기적으로 전송하여 정부 무결성 지표 데이터를 집계하고 통합해야 합니다.

### ② 평가 지표의 예

다음은 데이터 수집 단계별 평가 지표의 몇 가지 예입니다.



데이터 수집 단계

#### ○ 프로세스 단계

비즈니스 등록과 관련된 시간, 단계, 수수료 / 공급 업체의 송장 지불 속도와 정확성 / 일상적인 정보, 여권 또는 문서를 얻는 데 필요한 시

간, 단계, 요금 / 세금 평가, 징수, 그리고 기타 잡수익의 변동 및 예외의 시간, 단계 및 빈도, 범위 / 비즈니스의 일상적인 감사의 수, 빈도, 품질, 일관성 / 공공 조달에서 비경쟁 입찰을 사용하는 경우의 예외 빈도 및 추세 등.

○ 입출력 단계

입금된 자금의 처리 속도와 정확성 / 대행사가 부여한 보조금 또는 수당 지급 횟수의 추세 / 기본 서비스에 대한 요금 / 공공 서비스와 관련된 기본 물품에 대해 지불된 가격 / 상품의 수량 및 품질, 상품 수령 / 달성된 작업에 비례한 직원 수준 / 달성된 작업에 비례한 예산 수준 등

○ 성과 단계

공무원을 상대하는 비즈니스맨에 의해 발생하는 시간 또는 비율 및 관련 비용.

나) 온라인 여론 자료

① 평가 지표의 예

인터넷 시대가 도래함에 따라 점점 더 많은 사람들이 인터넷 플랫폼(예 : Facebook, Twitter, BBS 등)을 통해 정부 문제에 대한 의견과 제안을 표명합니다. 언론은 종종 정부 무결성 사건에 대해 공개하고 논평합니다. 전통적인 표본 조사에서의 '작은 표본' 방법과 달리, 인터넷에 대한 여론의 '대량 표본'에서 발굴된 정부 무결성 평가는 더욱 신뢰할 만합니다. 중요한 평가 지표에는 여러 정부 / 정부 기관 / 공무원의 네트워크 여론 사건 수, 부정 리뷰 / 긍정적 의견 수, 비율, 월별 / 연간 통계 등이 포함됩니다.

② 기술적 과정

정부의 무결성과 관련된 웹 사이트가 너무 많으므로 먼저 분산 크롤러 시스템을 구축해야 합니다. 대부분의 인터넷 여론은 텍스트 데이터이므로 자연어 처리는 여론 텍스트 분석에 적합합니다.

- 의미론적 이해와 텍스트 주제 검색 기술을 통해 정부 무결성과 관련된 여론을 선별해야 합니다.
- 자연어 처리를 통해 평가 대상을 추출하고 여러 정부 기관 또는 공무원의 불만을 선별합니다.
- 감정 분석을 통해 여론 평가의 질, 감정의 강도 등을 분석할 수 있습니다.

#### 다) 불만 신고 데이터

##### ① 평가 지표의 예

불만보고 및 그 처분 결과는 대중의 정부 무결성 평가와 밀접한 관련이 있습니다. 여러 정부 기관에는 자체 불만 제기 채널이 있습니다. 불만의 대상에 의한 불만보고 데이터는 두 가지 범주로 나눌 수 있습니다.

하나는 정부 부서 또는 공무원에 대한 것입니다. 이러한 종류의 데이터는 정부 무결성 평가에 직접 사용될 수 있으며, 보고서 수와 내부 고발자의 감정 강도를 고려할 수 있습니다. 다른 하나는 기업, 개인 또는 사회 단체에 대해 불평입니다. 이러한 보고서를 위해 정부는 조치 결과를 추적해야 합니다. 문제가 제대로 처리되지 않으면 불만 및 보고서가 정부의 문제로 바뀔 수 있습니다. 그렇다면 조치 만족도는 적절한 지표가 되어야 합니다. 또한 월별 / 연간 변경 사항이 있는 두 가지 유형의 불만보고 경향과 각기 다른 비즈니스 영역에서의 비율도 중요한 지표입니다.

##### ② 기술적 과정

기술 구현 측면에서 불만 및 보고된 데이터의 특정 조건에 따라 가장 경제적인 기술 솔루션을 포괄적으로 선택해야 합니다. 고도로 구조화된 도메인의 경우 데이터의 통계 분석을 직접 수행하는 것이 합리적입니다. 그러나 일부 도메인의 불만보고 데이터는 주로 텍스트 및 오디오 형식으로 저장되므로 다음과 같은 일부 빅데이터 기술로 처리해야 합니다.

- 음성 인식 기술을 통해 전화 녹음을 텍스트로 변환합니다.
- 자연어 처리를 통해 불만의 대상을 추출하고 XGBoost, SVM 등의 분류 기술을 통해 이러한 대상을 분류합니다.
- 감성 분석을 통해 정보 감각의 강도를 식별합니다. 특정 데이터 상황에 따라 특정 기술을 최적화해야 합니다.

## 5) 정부 무결성 평가 방법

다양한 채널을 통해 대량의 지표 데이터를 수집한 후 지표 데이터를 처리할 방법을 선택합니다.

위에서 언급한 바와 같이, 빅데이터를 기반으로 한 정부 무결성 평가의 목적은 평가 지수를 얻는 것뿐만 아니라 객관적이고 정량적인 데이터를 수집하는 것입니다. 과학적 방법을 통해 합리적이고 명확하며 해석 가능한 평가 표준 시스템이 정부의 특정 작업을 안내할 수 있게 공식화되었습니다. 기존의 평가 방법에는 퍼지 종합 평가, AHP 등이 포함됩니다. 이러한 방법은 평가 지표가 적고 질적 데이터가 적은 상황에 적합합니다. 학계에서는 SVM, GBDT, 신경망 및 기타 방법을 사용하여 무결성 등급을 탐색하고 있습니다.

## 6) 논의 사항

### 가) 응용의 어려움

데이터 수집은 빅데이터를 기반으로 한 정부 무결성 평가의 기초이자 어려운 지점으로, 초기 단계에서 공무원 간의 우려와 갈등을 유발하기 쉽습니다. 따라서 두 가지 중요한 점이 있습니다. 첫째, 정부 무결성 평가는 강력한 추진 부서에서 추진해야 합니다. 둘째, 표준화된 데이터 수집 및 평가 메커니즘이 확립되어야 하고 무결성 관리의 효과는 공무원의 성과 평가와 연계되어야 합니다. 또한 보상과 처벌 메커니즘이 있어야 합니다.

### 나) 정량적 및 정성적 응용의 조합

일부 시나리오에서, 수집된 정량적 지표는 제한될 수 있습니다. 따라서 양적 평가와 질적 평가의 조합을 고려할 수 있습니다. 그러나 정제된 수학적 모델의 사용은 정확한 평가 결과를 얻고 평가 결과의 객관성과 신뢰성을



높이는 데 도움이 되기 때문에 가능한 한 많은 양의 지표를 사용하는 것이 좋습니다.

#### 다) 데이터의 직접 적용

주관적 인식 데이터보다는 수집된 대량의 객관적인 정량적 데이터는 정부의 의사 결정 및 개혁을 더 잘 지원할 수 있습니다. 데이터는 벤치 마크와 함께 사용해야 합니다. 데이터 분석의 관점에서 적절한 벤치 마크는 지표의 평균 또는 중앙값일 수 있습니다. 데이터와 벤치 마크를 사용할 수 있게 되면 데이터 자체에 어떤 조치를 취해야 하는지 알 수가 있습니다.

- 각 도메인의 데이터를 자체 벤치 마크 세트와 비교 : 이러한 비교가 없으면 정책 결정자와 관리자는 결과가 좋은지 나쁜지 또는 문제의 규모가 상대적으로 크지 않은지 알 수 없습니다. 또한 벤치 마크와의 비교는 예를 들어 프로세스에 너무 많은 예외가 포함되어 있는지, 너무 빠르거나 너무 느리거나, 단계가 너무 많거나 너무 적거나, 지불된 가격이 너무 낮거나, 휘발성이거나, 너무 높은지 등의 여부를 판단하는 데 필수적입니다.
- 시간에 따른 데이터 비교 : 시간이 지남에 따라 결과가 향상되거나 저하되는지, 무결성 구축 전략에 즉각적인 이점이 있는지, 그리고 일부 데이터를 다른 방식으로 수집하거나 교체해야 하는지 등을 판단합니다.
- 한 도메인과 다른 도메인의 비교 : 전체 전략이 점차 구현됨에 따라 다양한 도메인 간의 비교가 가능해 지며 한 영역의 개선이 다른 영역에서 긍정적인 결과를 낳는지의 여부를 판단해 볼 수 있습니다.

#### 라) 정부 무결성 평가의 적용

빅데이터를 기반으로 한 정부 무결성 평가는 보다 정확한 평가 결과를 얻을 수 있을 뿐만 아니라 보다 신뢰할 수 있고 정부 부서의 공무원의 관행을 보다 잘 안내 할 수 있는 합리적이고 명확하며 설명 가능한 평가 표준 시스템을 개발할 수 있습니다. 평가 결과는 정부 또는 여러 지역의 여러 자매 기관과 비교할 수 있어 지방 정부가 문제의 핵심을 파악하는 데 도움이

됩니다. 정부 무결성에 대한 지속적인 평가를 통해 정부는 무결성 시스템과 메커니즘의 이행 효과를 추적할 수 있습니다. 평가 결과를 발표하면 정부 기관에 여론을 통한 압력을 가할 수 있습니다. 정부 부서에 대한 포괄적이고 과학적이며 정확한 무결성 평가는 지방 정부 성과 평가, 인사 선택, 정부 차입 및 기타 활동의 중요한 기초가 될 수 있으므로 광범위한 사회적 영향과 지방 정부와 공무원에 대한 심리적 압력 및 억제 효과를 형성합니다. 이러한 심리적 압력과 억제력의 도움으로 정부와 그 실무진은 불가피하게 무결성 문제에 대해 신중을 기하고, 점차 신뢰를 상실하는 것을 막고, 정부에 대한 무결성의 좋은 이미지를 확립할 것입니다.

## V. 결 론

모바일 기술, 소셜 미디어 및 ICT 활동이 증가함에 따라 공공 부문의 빅데이터 분석은 자연스러운 결과입니다. 공공 기관과 공공 정책을 혁신하기 위해 빅데이터 분석을 사용하는 정부 기관이 늘어나면서 경영진은 빅데이터에서 얻을 수 있는 통찰력은 무엇인지, 데이터에 대해 어떤 분석이 수행 되었는지, 그리고 결과에 대해 얼마나 확신하는지에 대한 의문을 던져야 합니다. 또한 빅데이터 분석에 사용되는 기술이 발전함에 따라 다양한 각도에서 데이터를 볼 수 있는 더 많은 옵션이 등장할 것입니다. 정부는 빅데이터 및 분석에서 나오는 새로운 정보를 사용하여 운영 비용을 절감할 수 있으며 결과를 개선하기 위해 민간 부문의 모범 사례를 채택해야 합니다. 정부에서 빅 데이터의 혁신적 가치를 더욱 높이려면 적절한 리소스 할당도 이루어져 데이터를 올바르게 캡처하고 관리해야 합니다.

빅데이터는 세계에 중요한 기회를 제공하며 정부 관리에 큰 관심을 불러 일으켰습니다. 데이터 리소스를 효과적으로 관리 및 사용하고 빅데이터의 깊은 가치를 발굴하면 정부 관리 및 의사 결정에 도움이 되며 조직과 정부에 특정한 방식으로 혜택을 줍니다. 조직에서 처리 및 분석되는 방대한 양의 데이터는 더 이상

빅데이터를 구현하는 데 있어 큰 문제가 되지 않습니다. 그러나 정부는 프로젝트에 대한 투자가 국가에 가치가 있도록 하기 위해 빅데이터 분석을 구현하기 위한 이니셔티브를 취하기 전에 효과적으로 계획을 세워야 할 것입니다.

빅데이터는 아직 초기 단계이지만 공공 부문의 거버넌스에 실제로 큰 잠재력을 가지고 있음이 밝혀졌습니다. 정부 기관은 빅데이터 분석을 적용하여 많은 양의 데이터를 보유함으로써 공공 서비스 제공, 일상적인 운영 방식, 정책 결정 및 기타 부가가치 서비스를 쉽게 개선할 수 있습니다. 그러나 시민의 사생활은 위험, 위협 및 도전을 최소화하기 위해 엄격하고 우선적으로 유지되어야 합니다. 또한 공공 부문 기관은 데이터의 부족, 불확실성, 행정가 및 정책 입안자의 효율성 부족으로 인해 이 기술을 채택할 충분한 준비가 되어 있지 않습니다. 모든 정부 기관은 부패, 위협 및 어려움을 줄이고 효율성, 책임 및 투명성을 높이기 위해 빅데이터 기술을 채택하여 투명하고 책임 있고 번거로움이 없는 정책을 펼쳐야 할 것입니다.

## 참고자료

1. Sherry L. Xie. (2018). Looking into Big Data: The Case of the U. S. Federal Government. KMIS (Knowledge Management Information System) 2018. DOI:10.5220/0006919001420150.
2. Effy Zalfiana Rusfian et al. (2018). Managing Government Digital Reputation through Big Data Processing. JKAP (Jurnal Kebijakan dan Administrasi Publik) Vol.22 (2), November 2018, 72-82.
3. Noor Zalina Binti Zainal et al. (2016). Big Data Initiatives by Governments - Issues and Challenges : A Review. 2016 6th International Conference on Information and Communication Technology for The Muslim World (ICT4M), November 2016, 22-24. DOI: 10.1109/ICT4M.2016.068.
4. Lingdan Liao. (2017). Grid Information Management Innovation of Government in the Era of Big Data. 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC), October 2017, 3-5. DOI: 10.1109/ITOEC.2017.8122519.
5. Xue Sixin. (2017). A New Governance Architecture for Government Information Resources based on Big Data Ecological Environment in China. 2017 IEEE International Symposium on Multimedia (ISM), December 2017, 11-13. DOI: 10.1109/ISM.2017.103

6. Zheng Yan. (2018). Big Data and Government Governance. 2018 International Conference on Information Management and Processing (ICIMP), January 2018, 12-14. DOI: 10.1109/ICIMP1.2018.8325850.
7. Md Nazirul Islam Sarker et al. (2018). Smart Governance through Bigdata: Digital Transformation of Public Agencies. 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD), May 2018, 26-28. DOI: 10.1109/ICAIBD.2018.8396168.
8. Katsuyuki Yamazaki et al. (2018). Governments in the Age of Big Data and Smart Cities. IEEE (Institute of Electrical and Electronics Engineers) COMPUTER SOCIETY. DOI: 10.1109/MC.20.
9. Ping Yang et al. (2019). Research on Government Integrity Evaluation Based on Big Data. 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD), May 2019, 25-28. DOI: 10.1109/ICAIBD.2019.8837005.
10. Karl L ofgren et al. (2020). The value of Big Data in government: The case of 'smart cities'. Big Data & Society January - June 2020: 1 - 14. DOI: 10.1177/2053951720912775.
11. Chad Squitieri. (2015). CONFRONTING BIG DATA: APPLYING THE CONFRONTATION CLAUSE TO GOVERNMENT DATA

COLLECTION. 101 Va. L. Rev. 2011 (2015). Available at SSRN: <https://ssrn.com/abstract=3015056>.

12. Pwint Phyu Khine et al. (2017). Big Data in Government Organizations: A Review. 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), March 2017, 10-12. DOI: 10.1109/ICBDA.2017.8078855.

13. Rhoda C. Joseph et al. (2013). Big Data and Transformational Government. IT Professional, VOL 15, Nov-Dec 2013, 43-48. DOI: 10.1109/MITP.2013.61.

14. British and Irish Association of Law Librarians. (2014). Big Data and the Implications for Government. Legal Information Management, 14, 253 - 257. DOI:10.1017/S1472669614000553

15. Zaher Ali Al-Sai et al. (2017). Big Data and E-government: A review. 2017 8th International Conference on Information Technology (ICIT), May 2017, 17-18. DOI: 10.1109/ICITECH.2017.8080062.