

RGT 및 Q방법론을 활용한 빅데이터 거버넌스의 주요 구성요소 개발

강 지 연*

한양대학교 교육공학과

본 연구의 목적은 효율적이고 효과적인 빅데이터 거버넌스 구축을 위해 빅데이터 거버넌스의 개념 정의 및 주요 구성요소를 도출하는 데 있다. 이를 위해 이론연구와 현장연구를 구분하여 선행연구, 문헌 및 사례분석을 통해 기초자료를 수집하고 이를 기반으로 질적 및 양적 연구방법이 혼합된 RGT기법과 Q방법론을 순차적으로 적용, 연구 결과를 도출하였다. 연구 대상은 빅데이터 관련 직무경력을 기준으로 현장실무자와 전문가로 구분하여 단계적으로 연구를 수행하였다. 연구 결과 첫째, 빅데이터 거버넌스를 '빅데이터 관련 전략 수립과 목표달성 및 가치 창출을 위한 협업 및 통합관리체제'로 정의할 수 있었고, 둘째, 주요 구성요소로는 '정책', '조직관리', '프로세스', '데이터 분석 및 관리' 등으로 도출되었으며, 셋째, 빅데이터 거버넌스를 인식하는 유형은 '정책 및 조직관리 중시형', '성과지향 프로세스 중시형', '데이터 분석 및 관리 중시형'의 세 가지로 구분되었다. 본 연구는 이론 및 실증연구가 순차적으로 수행되었고, 질적 및 양적 연구방법을 동시에 적용하였다는데서 연구의 학문적 기여를 찾을 수 있으며, 본 연구 결과를 빅데이터 거버넌스의 구축 및 빅데이터 생태계 연구의 기초자료로 활용 가능할 것이라는 데서 실용적 의의를 찾을 수 있다.

주요어 : 빅데이터, 빅데이터 거버넌스, RGT기법, Q방법론

* 주저자: 강지연/한양대학교 교육공학과 박사수료
E-mail: edukang1203@gmail.com

I. 서론

4차 산업혁명에 따른 지능과 정보에서의 기술혁신을 바탕으로 실시간으로 끊임없이 증가하는 다양한 형태의 데이터가 생산되고 있다. 이는 빅데이터의 개념으로 발전하여 고차원적인 분석방법론의 적용으로 유의미한 가치를 발견케 하고 있다(Daniel, 2019).

국내에서도 정부 차원의 한국판 뉴딜종합계획을 발표하며 디지털 뉴딜 분야로 'DNAData, Network, AI 생태계 강화'를 선두로 한 4개의 분야가 빅데이터의 고도화된 활용을 목표로 설정하였다. 또한 기업 및 기관 등의 다양한 분야에서도 스마트융합시대의 주요 패러다임을 선도하기 위해 빅데이터 활용을 핵심 주력사업으로 삼고 있어 빅데이터의 활용이 경쟁력 확보의 중추역할을 할 것으로 전망된다(정부부처합동, 2021).

특히 Covid-19 이후 교육, 의료분야를 필두로 비대면 디지털 인프라를 통한 다양한 데이터가 폭발적으로 증가하게 되면서, 머신러닝, 딥러닝 등 최신 빅데이터 분석기법을 활용한 관련 연구를 통해 학습환경 개선, 서비스의 질 향상, 정책 결정 등으로 부가가치를 창출하고 있다.

하지만 이러한 잠재력과 고부가가치에도 불구하고 현재 빅데이터 관련 조직에서의 효율적인 전략 및 조건에 대한 인식은 여전히 부족한 실정이다. 빅데이터 개념의 도입 시 데이터 수집과 처리 등 IT 기술 적용에만 집중하여 빅데이터 활용에 대한 체계적인 프로세스 및 관리가 제대로 설계되지 못했고, 이로 인해 관련 조직 간의 사일로현상(silo effect), R&R 모호, 디지털 리더십 부재 등 미성숙한 협업 및 관리로 축적된 빅데이터들이 유의미하게 사용되지 못한 채 소멸되고 있다(Soares, 2018; Daniel & Butson, 2013).

이러한 맥락에서 빅데이터의 효과적인 활용을 위한 핵심요인으로 빅데이터 거버넌스(Big data gov-

ernance)가 조직의 가치창출 및 성과달성을 위한 필수조건으로 인정받고 있다(Soares, 2018, 2013; Kelle, 2015; SAS, 2014). 글로벌 리서치기업 M&M은 전 세계 데이터 거버넌스 규모가 2020년 21억달러에서 5년 내 57억 달러로 증폭될 것으로 예측하였으며, 데이터 관리, 비즈니스 협업 증가, 규제 및 준수사항 증가 등의 요소가 빅데이터의 이니셔티브를 주도할 것으로 전망하였다(M&M, 2020). 빅데이터 활용을 위해 효율적이고 효과적으로 빅데이터를 관리, 운영할 수 있는 체제로서의 거버넌스에 대한 고찰과 연구가 요구되는 상황이다(Roberts, Chang, & Gibson, 2017).

빅데이터 거버넌스 구축은 단순히 데이터를 잘 분석하기 위한 하나의 절차 항목이 아니며, 통찰력 있는 분석전문가와 비즈니스 사용자 및 경영진이 가치 있고 올바른 질문을 던지고, 패턴을 인식하며, 정보에 근거한 가정을 세우고, 행동을 예측하는 등의 전체적인 발견 체계라 할 수 있다(Oracle, 2021).

빅데이터 거버넌스를 구축하는 목적은 빅데이터 시스템의 구조와 범위 및 영역을 설정, 디자인하여 효율적인 분석 프로세스를 통해 조직의 궁극적인 목표와 가치를 달성하는 것이며, 나아가 현재 빅데이터 거버넌스의 수준을 진단, 향상하기 위한 가이드를 제공해주는 데 있다(Soares, 2013). 이러한 빅데이터 거버넌스를 구축하기 위해서는 빅데이터 관련 개념 정의 및 구성요소 도출이 선행되어야 할 것이다.

기존의 데이터 거버넌스 관련 연구들은 대부분 데이터를 분석하기 위한 IT기술 및 데이터 품질관리에만 초점을 맞추거나 데이터 관리의 프로세스상의 문제를 해결하기 위한 프로젝트성 연구를 수행해왔고, 실제 빅데이터 거버넌스의 구성요소에 대한 실증연구가 부족한 상황이다(박정, 조완섭, 2020; 송경재, 장우영, 조인호, 2018; 최영환, 조완섭, 이경희, 2018; 조완섭, 2018; 장경애, 2017; Soares, 2018). 또한 기존의 이론연구 외에 실제 현장에서의 목소리

를 심도있게 분석한 연구는 이루어지지 않고 있다 (권숙진, 한재훈, 신중호, 2017; Daniel, 2019).

빅데이터 거버넌스는 정책학, 심리학, 분석학, 교육공학 등 다양한 분야의 이론이 융합적, 복합적으로 적용된다. 다양하고 적절한 이론에 심층적인 현장 및 실증연구를 더해 개념 정의와 주요 구성요소가 도출되어야 할 것이다. 지금까지 서술한 내용을 종합하여 다음의 연구문제들을 도출하였다.

- 연구문제 1: 빅데이터 거버넌스의 개념 정의는 어떠한가?
- 연구문제 2: 빅데이터 거버넌스의 주요 구성요소는 무엇인가?
- 연구문제 3: 빅데이터 거버넌스에 대한 인식유형은 어떠한가?

본 연구에서는 빅데이터 거버넌스를 구축하기 위한 관련 개념정의와 주요 구성요소를 도출하는 것을 목적으로 하며, 이를 기반으로 빅데이터 거버넌스의 실증적 모델 수립에 기초자료로 활용될 수 있을 것으로 판단된다.

본 연구의 구성은 다음과 같다. 2장에서는 연구의 이론적 배경으로 개념 및 연구 방법론을 소개하고, 3장에서는 연구의 설계 및 절차, 분석방법을 설명한다. 4장에서는 실험결과, 마지막 5장에서는 논의 및 시사점, 향후 연구과제를 제안, 논의한다.

II. 이론적 배경

1. 빅데이터 거버넌스

빅데이터(Big data)는 다양하고 빠르게 변화하며 실시간으로 축적되는 대규모 데이터들을 수집하고 분석하는 과정에서 나오는 다양한 이슈를 다룬다. 빅데이터는 규모(volume), 다양성(variety), 속도

(velocity)라는 3가지 측면의 ‘3V’로 설명하는 것이 가장 일반적이다(Laney, 2001). McKinsey Global Institute(2011)는 ‘일반적인 데이터 처리 방식인 데이터의 수집, 저장, 관리, 분석 범주를 증가하는 규모의 데이터셋’이라고 데이터 규모에 초점을 맞추어 빅데이터를 정의하였고, Gartner(2011)는 ‘대용량(high-volume), 다양성(high-variety), 빠른 속도(high-velocity)’라는 특성으로 설명하였다. 최근에는 데이터 신뢰성(veracity), 데이터 분석 결과의 시각화(visualization), 조직에 필요한 통찰, 성과를 창출할 수 있게 하는 데이터의 가치(value) 등의 중요한 특성을 추가하여 ‘6V’로 설명하기도 한다(국립중앙과학관, 2020; Oracle, 2021).

최근 선진 국가들이 빅데이터를 조직의 중요한 자산으로 간주하여 체계적인 관리에 집중 투자하고 있지만, 국내 상황은 아직 HW/SW만 잘 구비해놓으면 데이터의 가치가 자연히 높아질 것이라고 기대하는 경우가 많다(조완섭, 2018). 하지만 빅데이터를 가치있게 활용하기 위해서는 단순히 HW/SW를 정비하는 것으로는 충분하지 않다. 전사 차원의 데이터베이스 구축 및 활용뿐만 아니라 빅데이터 거버넌스 구축을 통해 효과적이면서도 완전한 데이터 관리체계를 실현해야 한다(조완섭, 2020; Oracle, 2021; Soares, 2018).

거버넌스의 어원은 ‘(배의) 키를 잡다’는 뜻의 라틴어 ‘gubernare’에서 유래했는데(Muller, 2014), 이후 국정관리체계, 공공관리, 통치양식, 국가경영, 협력적 통치, 네트워크적 관리, 협치 등 여러 가지 의미로 번역되어 왔다(이정규, 정승렬, 2019). 현재 가장 일반적인 개념으로는 ‘거버넌스는 정치과정 방식으로, 정치행위자들이 특정 목표를 가지고 수평적인 네트워크(horizon network)를 구성하는 자율적 협의와 조정의 과정’이며, 다른 다양한 개념 정의의 대부분 거버넌스를 조직의 ‘공동문제 해결’과 관련된 ‘의사결정 과정’을 광범위하게 포괄한 개념으로 정의하고 있다(이정규, 정승렬, 2019; 한국전산원,

2004).

거버넌스는 정치·행정학을 비롯한 사회과학 분야에서 연구가 활발히 진행되고 있으며, 현재 거버넌스 이론이 주목받는 가장 큰 이유는 21세기 정보화·세계화로의 급격한 사회변화에 대응하기 위한 국가중심 정치로부터의 변화라고 할 수 있다(송경재, 장우영, 조인호, 2018). 21세기에는 네트워크 거버넌스로서 국가가 통치 주체로서의 주도적인 역할을 강조하기보다는 사회가 문제해결의 통치능력을 가지고 있다고 가정하고 다양한 행위자들과의 수평적 네트워크에서 대등하게 참여하는 방식에 의해 이루어진다는 점을 강조하는 것이다(현영란, 2015).

<표 1> 데이터 거버넌스에 대한 다양한 정의

정의	연구자
전사 차원의 전체 데이터에 대해 정책, 지침, 표준화, 운영조직 및 책임 등의 표준화된 관리체계를 수립하고, 이를 운영하기 위한 프레임워크를 구축하는 것	한국데이터베이스진흥원 (2019)
조직의 데이터 활용, 관리능력이 효과적으로 이루어지도록 평가, 감독, 모니터링(EDM)하는 체제	국제표준화기구 (2017)
전사 차원에서 보유하고 있는 모든 데이터에 대하여 정책·지침·표준·전략 및 방향을 수립하고, 데이터를 관리할 수 있는 조직과 서비스를 구축하는 관리시스템	조완섭 (2017)
데이터 자산관리에 대한 권한·계획·통제로, 이를 위한 모니터링이자 집행과 관련된되는 공동의 의사결정	Ladley (2012)
데이터 경영, 분배, 보호를 위한 프레임워크 및 로드맵	김석수 (2011)
기업이 데이터를 전사적 자산으로 활용 가능하도록 인력(people), 프로세스(process), 기술(technology)을 조정, 조율하는 활동	Soares (2010)

데이터 거버넌스는 ‘전사 차원의 전체 데이터에

대하여 지침 및 정책, 표준화, 운영조직 및 책임 등의 표준화된 관리시스템을 수립하고 운영을 위한 프레임워크를 구축하는 것’(한국데이터베이스진흥원, 2019), ‘조직의 데이터 활용 및 관리능력이 효과적으로 이루어지도록 평가·감독·모니터링(EDM: Evaluation, Direction, and Monitoring)하는 체제’(국제표준화기구, 2017) 등으로 설명된다. 데이터 거버넌스에 대한 다양한 정의를 정리해보면 <표 1>과 같다.

이러한 데이터 거버넌스는 특정 프로세스 및 시스템을 목표로 하기 보다는 전사차원의 모든 정형화된 데이터의 보안 및 안정성, 품질관리에 중점을 둔다(Kim & Cho, 2018). 하지만 빅데이터는 비정형 및 반정형 데이터뿐만 아니라 실시간으로 축적되는 고용량 데이터의 저장 및 처리를 필요로 하며, 로그 정보와 같은 정밀하고 미세한 데이터를 다루어야 하기 때문에 일반적인 데이터와 다른 별도의 접근방법이 필요하다(Edjlali & Friedman, 2011). 특히 사회적 동향 분석 및 변화 예측과 같은 광범위한 빅데이터 분석과 활용을 통해 빅데이터의 핵심 역할인 ‘value’가 창출되어야 하며, 이를 위해서는 조직의 비전과 전략을 반영할 수 있는 빅데이터 거버넌스의 프레임워크가 요구된다.

빅데이터 거버넌스의 궁극적인 목적은 우수한 품질의 데이터 확보와 체계적 관리, 적극적 활용을 통해 조직이 추구하는 다양한 가치 창출에 지속적으로 기여하는 데 있다(최영환, 조완섭, 이경희, 2018). 또한 빅데이터 관리를 위한 비용 절감과 활용도 확대를 통해 개별데이터의 가치를 향상시키며, 고품질 데이터가 조직의 비즈니스 목적에 부합하고 최적화된 서비스를 지속적으로 제공 가능하도록 효과적인 관리를 하는 데 있다(조완섭, 2017).

실시간으로 쏟아지는 빅데이터 분석, 예측 결과로부터 미래에 관한 통찰을 얻고, 중대한 의사결정에 활용하기 위해 빅데이터를 목적과 시간에 맞게 통합, 정제, 분석하여 활용하려면 이를 위한 적절한

관리체계 즉, 빅데이터 거버넌스 없이는 불가능하다 (Edjlali & Friedman, 2011).

하지만 빅데이터 거버넌스라는 구체적인 주제에 대한 선행연구는 많지 않다. 대부분의 선행연구들이 빅데이터를 전략적, 행정적 시스템 차원의 거버넌스로 논의를 발전시키기보다는 주로 빅데이터 분석, 클라우드, AI, 데이터 알고리즘 및 아키텍처 등 빅데이터와 관련된 새로운 기술이나 인프라를 적용하기 위한 프로세스를 제안하거나 빅데이터가 가지고 있는 문제점 및 경고를 다루는 정도이다(Ali, Ali, & Asharul, 2018; Akoka, Comyn-Wattiau, & Laoufi, 2017). 최근의 연구 역시 정보사회에서의 빅데이터의 위험성을 경고하거나, 프라이버시 보호 문제, 디지털 시민의 등장과 이를 둘러싼 정보주체의 문제, 국가차원에서의 통치적 거버넌스 구성 등 이론적 탐색 및 제안에 그치고 있다(박정, 조완섭, 2020; 송경재 등, 2018; 최영환 등, 2018; 김지연, 2015).

이러한 연구의 연장선상으로 효율적이고 효과적으로 빅데이터를 관리, 운영할 수 있는 체계로서의 거버넌스에 대한 고찰과 연구가 요구된다. 빅데이터를 분석, 활용하여 조직이 혁신되고 성과 달성 및 효율적인 정책결정이 가능하기 위해서는 빅데이터 거버넌스에 대한 장기적인 관점에서의 체계적이고 단계적인 연구가 필요하다.

2. 레퍼토리 그리드 기법

레퍼토리 그리드 기법(Repertory Grid Technique, RGT)은 인지심리학자인 George Kelly(1955)에 의해 제안된 구성심리학 기반의 연구방법론으로, 개인의 경험에 의한 인지구조를 분석하는 구조화된 심층 인터뷰 기법이다.

인간은 자신의 경험이 해석되는 의미 체계를 스스로 창조해내는데, 이 과정을 심층적으로 탐구하는 접근이 구성주의 심리학이다(Raskin, 2002). 구성심리학 연구는 세계 여러 나라에서 꾸준히 연구되고

있으며, 구성주의 연구의 대다수가 RGT를 사용할 정도로 광범위한 활용성을 갖는다(김소희, 유금란, 2018; Saul, et al., 2012), 국내에서는 RGT기법의 활용도가 미미하다.

RGT의 기본적인 전제는 객관적 진실 및 절대적 진리란 존재하지 않으며, 세계는 개인 자신이 해석하는 방식으로 존재한다는 것이다(Kelly, 1955). ‘해석하다(construe)’의 어원은 라틴어 ‘construere’인데, 이는 ‘구조(structure)를 배열(arrange)하고 형태를 부여한다(form-give)’라는 의미를 내포한다(Mahoney, 2003). 즉, ‘해석한다’라는 것은 개인의 구조에 의해 의미체계가 구성되는 것을 뜻한다.

개인의 인지구조 혹은 구성체계(construct system)는 개인의 역사적 산물이자 삶의 경험이 누적되어 형성된 ‘관점’으로, 삶의 사건들을 구별하고 통합하며 예측하게 도와준다. 구성개념(construct)은 독단적인 개체가 아니라, 수많은 다른 구성개념들과 서로 연계되어 있으며, 이러한 연결을 통해 복잡한 의미 체계가 생성된다. 이를 개인구성체계(personal construct system)라고 하며, 개인이 가지고 있는 이론(theory)과 동일하다(Hardison & Neimeyer, 2007).

개인구성체계의 탐색을 통해 ‘단어 너머의 것’을 알 수 있게 되고, ‘맥락’을 연구할 수 있게 되며, 개인의 언어세계를 이해하여 사전이 주지 못하는 개인의 견해를 이해할 수 있게 된다(김소희, 유금란, 2018). 현실은 객관적이고 보편적으로 존재하는 것이 아니라 사회적으로 구축되거나 개인의 상호작용에 의한 주관적인 구성체라고 보는 것이다(McNamee & Gergen, 2004). 이처럼 외부와의 상호작용을 통해 외부의 객체에 대한 견해 및 인식이 발전되는데, 객체에 관한 이러한 개념들은 지금까지 경험한 다른 객체와의 유사함과 차이들의 조합으로 구성되며, 이는 RGT를 통해 추출될 수 있다(Kelly, 1955).

수많은 연구자들에 의해 실제로 개인이 내부적으로 인지구조(construct)를 이용하여 외부 환경을 표현하며, RGT가 이러한 인지구조를 추출하는데 효과

적이라는 사실을 입증했다(Mair, 1966). 그래서 지금까지 심리상담을 비롯하여 마케팅, 정책 등 다양한 분야에서 활용되고 있는데, RGT를 독립적으로 연구에 사용하거나, 정성적 및 정량적 연구조사를 위한 예비분석에 활용하기도 하며, 연구에서 얻은 결과의 유효성을 검증하거나 심화하기 위한 보완기법으로 사용되기도 한다. 양적 연구에서 제한적인 개인의 내적 현실을 탐색하고, 분석 과정에서의 비효율성과 자료 요약으로 인한 정보 손실이 발생하는 질적연구의 한계를 보완 가능하다는 점이 RGT의 연구가치가 인정받는 이유이다(황선정, 조성호, 2011).

RGT는 요소(elements), 구성개념(constructs), 평정(rating)의 세 가지 핵심요인으로 구성된다(Fransella et al., 2004).

요소는 해당주제에서 의미를 가지는 대상이나 경험으로, 사람, 사건, 사물 등의 형태로 참여자에게 제시된다. 질문과 함께 제시된 요소는 참여자의 인지구조에 따라 유사점과 차이점으로 구별되고, 나머지 요소들을 추가하며 반복된 질문을 통해 구별과 연결을 거쳐 구성개념이 도출된다.

구성개념은 요소에 대한 참여자의 양극화(bipolar)된 해석으로, 인지구조상에 양극성을 적용하여 구성개념을 도출할 수 있다. ‘도출된다’는 것은 인터뷰 과정을 통해 응답자의 구성개념이 언어로 표현되는 것을 의미한다. RGT의 최종목적인 구성개념을 도출하기 위해 요소를 제시하는 방법에는 한 번에 하나의 요소를 보여주는 모나딕(Monadic), 두 개의 요소를 동시에 보여주는 다이아딕(Dyadic), 세 개의 요소를 동시에 제시하는 트리아딕(Triadic)의 세 가지 방식이 있다. 가장 많이 사용되는 방식은 트리아딕으로, 세 개의 요소 중 유사한 것 2개와 나머지 1개로 분류하고, 그 기준에 의해 인지구조가 파악되고 구성개념을 추출할 수 있다. 구성개념은 ‘좋다-나쁘다’, ‘차갑다-뜨겁다’ 같은 양극(bipolar) 개념으로 도출되는데, 이를 척도화, 정량화하여 분석데이터로 활용한다. 요소와 구성개념

의 정해진 개수는 없지만, 참여자의 피로도 및 소요 시간, 구성개념의 중복 등을 고려하여 적정 개수는 10~20개 사이로 정하는 것이 권고되고 있다(Fransella et al., 2004).

평정(rating)은 각 요소에 구성개념을 적용하여 점수를 부여하는 과정이다. 각 요소에 대해 구성개념이 가지는 값을 통상 5~7점 급간의 리커트척도로 평정하게 된다(Fransella, et al., 2004).

RGT의 참여대상에 대한 최소한의 인원제한은 없다. 최소 1명부터 25명 이상을 참여자로 구성한 연구까지 다양하다. 일반적으로 5~10명 정도면 RGT를 활용한 연구에서 중복없이 유의미한 개념적 지식을 적절히 추출할 수 있는 것으로 여겨진다(Crudge & Johnson, 2007; Dunn et al., 1986).

3. Q방법론

Q방법론은 개인이 가지고 있는 자아참조적 주관성(operant or self-referent subjectivity)을 분석하는 양적 및 질적 연구의 혼합방법론으로, 인간의 주관성 연구를 위하여 정치학, 정책학, 심리학 등 사회과학 전반에 걸쳐 활용되고 있는 접근 방법이자 분석 방법이다(길병옥, 2020; Stephenson, 1953).

Q방법론은 개인의 주관성 경향이나 가치관에 대한 정량화(Quantification of patterned subjectivities)를 통하여 해석된 가치유형을 분석하는 방법론으로서, Q방법론은 질적 연구인 동시에 양적 연구이며, 질적인 접근과 양적인 접근이 동반된 혼합 연구의 방식을 적용한다(Stenner, 2011; Ramlo & Newman, 2011). 이러한 방식에 대해 Stenner & Rogers(2004)는 ‘qualiquantology’라는 용어로 설명했다.

Q방법론을 처음 제안한 William Stephenson(1953)은 전통적인 요인분석(Factor Analysis)을 R연구로 명명하며 R요인분석의 대안으로 인간의 주관성에 대한 연구는 사람 간의 상관관계를 기본으로 하는

Q방법론이 더욱 적합하다고 하였으며, 인간의 주관성에 대한 양적연구의 측면, 즉 실증적 연구의 가능성을 보여주었다. R방법론에서 표본 간의 점수 차는 표본들이 가진 해당 변인 속성에서의 개인차(individual difference)가 반영된다는 이론에 의거한다. 반대로 Q방법론은 진술문(Q-statements) 간의 점수 차는 개인에게 있어서 진술문의 중요도의 차이를 반영하는 것이기 때문에 의미상의 개인 내 차이(intra-individual difference in significance)에 근거한다(Stephenson, 1953). 이러한 과정을 통해 참여자는 그들 스스로의 언어로서 모델링 해나가며 자신의 주관성 구조를 표현하게 된다(김순은, 2010).

양적 연구가 신뢰도와 타당도를 중시하고 질적연구가 신뢰성과 의존가능성 확립을 모색하면서 유사한 상황에서의 적용을 탐색하는 전사가능성을 중시하는 것과 달리, Q방법론은 신뢰도와 타당도 문제에 대해 상이하면서도 고유한 입장을 보유하고 있다(Bloomberg, & Savin-baden, 2010). Q표본의 확정 과정에서 Q분류를 반복 진행하는 방식의 재검사로 신뢰도 검증은 하고 있지만, 한 개인의 고유한 관점에 대해 외부적 준거가 간섭할 수 없기때문에 타당도는 무의미하다(김홍규, 2007; Brown, 1980). Q방법론이 조작적 정의가 아닌 개인 내의 의미의 차이와 심리적 위계가 측정 단위가 되는 자결적 반응에 근거하기 때문이다(김홍규, 2008).

Q방법론은 자아참조적인 주관적 인식의 집합체(concourse)를 대상으로 유형(factor)의 경험, 태도 등을 분석하는 방법론이다. 빅데이터 거버넌스는 빅데이터 생태계라는 복잡계가 거버넌스의 정치·행정학, 사회학, 분석학, 교육공학 등 이론과 융합, 적용되어 구성된 체제로, 이에 대한 인식은 인간의 내면적 경험과 태도를 포함한 주관성을 띠고 있다는 점에서 다양한 개체 및 프로세스의 협력으로 구성되는 빅데이터 거버넌스의 연구에 매우 적합하다고 볼 수 있다.

III. 연구방법

1. 연구 설계

본 연구는 빅데이터 거버넌스의 주요 구성요소를 도출하기 위해 아래와 같은 단계로 수행된다.



<그림 1> 연구 설계

데이터 수집을 위한 첫 번째 단계는 이론연구로, 선행연구 및 문헌분석, 사례조사를 통해 빅데이터 거버넌스의 주요 구성요소 도출을 위한 1차 요소를 선정하였다. 두 번째 단계는 현장연구로, 전 단계에서 선정된 1차 요소를 대상으로 브레인스토밍 및 FGI(Focus Group Interview)를 통해 최종 요소를 추출, 확정하였다.

데이터 수집 단계에서 정의된 개념과 추출된 요소를 대상으로 데이터 분석을 수행하게 되는데, 첫 번째 단계에서는 현장실무자를 참여대상으로 선정하고, RGT(Repertory Grid Theory) 기법을 통해 총 15개의 구성개념을 도출하였다. 다음 단계로 현장전문가를 대상으로 질적 및 양적 혼합연구 방식의 Q방법론을 적용하여 최종적으로 빅데이터 거버넌스의 주요 구성요소를 도출하였다.

2. 조사 및 분석방법

분석단계에서는 이론연구를 통해 논의된 관련 개념들의 개괄적 정의를 활용하여 현장실무자 및 의사결정자, 관련 전문가들을 대상으로 FGI, RGT, Q 방법론을 단계별로 적용, 분석하였다.

FGI에서는 미리 작성되고 표준화된 폐쇄형 질문으로 이루어진 구조적인 조사 인터뷰와 개방적이고 명백하게 비구조적이며 인류학적인 인터뷰 방식을 활용한다. 본 연구에서는 전 단계인 이론연구를 통해 도출된 개념 정의 및 요소를 대상으로 개방형과 폐쇄형 질문을 혼합한 반구조적 면접(semi-structured interview)을 활용하였다.

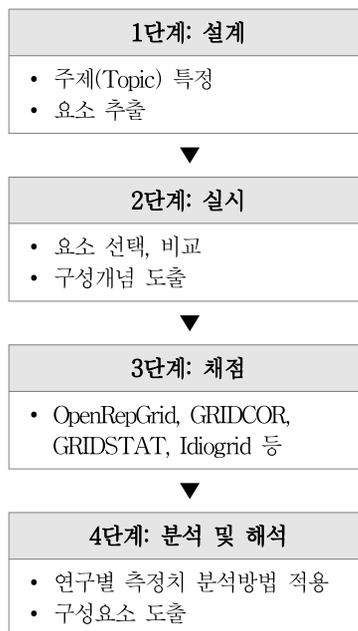
앞서 문헌 및 선행연구 등 이론연구를 통해 1차로 도출된 개념 정의를 기반으로 20개의 요소(element)를 대상으로 구조화된 질문을 작성하고, 질문 후 실제 답변에 따른 ‘why’와 ‘what’을 중심으로 비구조화된 질문을 심층적으로 이어나가는 반구조화된 방식의 FGI를 실시하였다.

개념 제정의 및 최종 요소 도출을 위해 FGI에 참여한 대상으로는 빅데이터 활용 프로세스 상에 직접 관련 실무를 담당하고 있는 3년 이상의 현장실무자 5명을 선정하였다. 표본선정은 목적표집법을 선택하였는데, 실제 관련 경험이 있는 연구자의 통찰에 의한 특정 대상 선정은 그 집단의 대체적인 문화와 구조를 파악하기 용이하여 질문과 분석의 질이 제고되는 장점이 있다(Weiss, 1994).

FGI를 통해 최종 정의된 개념 및 추출된 요소를 대상으로 RGT를 통해 구성개념(construct)을 도출하였다. 인터뷰 참여대상은 빅데이터 관련 3년 이상의 빅데이터 관련 현업종사자 8명으로 선정하였다.

RGT는 기본적으로 다음 4단계를 거쳐 구축, 분석된다.

<표 2> RGT의 4단계 프로세스



추출된 요소를 대상으로 참여자가 세 요소 중 유사한 두 개와 나머지 다른 한 개를 양분하는 과정에서 구성개념이 도출되는 트리아딕(triadic) 방식을 적용하였다. 분석 및 시각화 도구로 Idiogrid, OpenRepGrid, R 프로그램을 분석단계에 적합하게 선택하여 사용하였으며, 구성개념의 분석결과에 대해 직관적인 이해를 돕고자 Bertin Display, Biplot 등으로 시각화하였다.

RGT를 통해 도출된 구성개념을 바탕으로 Q진술문을 작성하였다. Q진술문 작성은 Q방법론 적용에 있어 가장 중요한 부분인데, 작성된 진술문에 의해 연구 결과가 해석되기 때문이다(김홍규, 2008).

본 연구에서는 문헌 자료와 같은 2차 자료와 심층인터뷰를 병행하여 수집된 구술형 및 추출형 혼합방식이며, 연구 설계 방법은 RGT 결과를 바탕으로 연구자가 사전에 설정한 기준에 따라 Q표본을 수집하는 구조적 표본추출을 적용하였다. Q진술문은 내용이 어렵고 미묘하거나 비슷한 내용으로 구성되거나 진술문의 개수가 너무 많으면 연구의 신

뢰도에 문제가 발생한다(김홍규, 2008). Q진술문은 간결하고 용이하게 작성되어야 하며, 통상 30~60개 정도의 진술문이 연구에 가장 적합한 것으로 제시된다(Brown, 1980; 김순은, 2010, 2007; 김홍규, 2008). 본 연구에서는 선행과정을 거쳐 각 구성개념당 5~6개의 진술문을 작성, 총 78개의 진술문을 모집단으로 구성하였다. 자의적 해석이 가능하거나 사실을 설명 및 묘사하는 진술문은 삭제하여 52개의 문항으로 정리하였으며, 내용타당도 검증을 위해 관련 전문가 3인(AI·빅데이터과 교수, 데이터과학과 교수, IT정보센터장)에 의해 전문가 내용타당도인 CVI점수를 산출하였다. Lynn(1986)이 제시한 내용타당도 지수(content validity index)의 기준에 의해 1.00 미만의 문항 7개를 삭제 및 수정하여 최종적으로 총 47개의 Q표본을 추출하였다.

Q방법론은 주관성에 의한 유형 분류 연구이므로 대표성과 연구의 질을 제고하기 위한 P표본의 선정이 매우 중요하다. Q방법론은 개인 간의 차이가 아니라 개인 내의 차이를 다루는 것이므로 P표본의 수에 아무런 제한을 받지 않으며, Stephenson은 한 명의 P표본으로 연구를 진행한 사례도 있었다(김홍규, 1990; 한국주관성학회, 2014). 적절한 표본수가 제시된 바는 없지만, 4명 정도면 충분하다는 의견(김홍규, 1998)과 최대 표본수가 50명을 넘지 않는 것이 좋다는 의견(김순은, 2007)이 지배적이다. 다만 빅데이터 거버넌스의 구성에 대한 심층적 평가는 오랜 현장 경험을 통한 고도의 전문성을 반영해야 하며, 연구주제와 관련하여 특정 성향의 표본으로 편향되어 구성되지 않아야 한다.

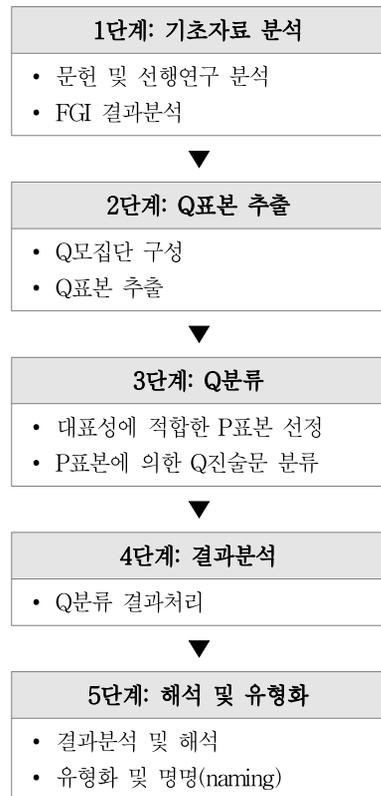
이에 따라 RGT를 통해 도출된 주성분과 연관된 조직 중심으로 빅데이터 관련 5년 이상 직무를 수행한 선임연구원 및 팀장급 이상의 의사결정권자 7인과 빅데이터 관련 박사학위를 소지한 5년 이상의 현업전문가 3인의 총 10인을 전문가 집단으로 구성하고 이들을 P표본으로 선정하였다.

Q방법론의 가장 일반적인 프로세스는 사전자료

분석 및 FGI, Q모집단 구성 및 Q표본 추출, P표본 선정 및 Q분류, 결과분석, 해석 및 유형화의 5단계로 진행된다. Q모집단 구성 및 추출단계에서는 FGI 결과와 문헌조사 등을 반영하여 Q모집단을 구성한 후 구성된 모집단 중 의미가 불명확하거나 중복된 질문을 제외하고 Q표본을 추출한다. P표본 선정과 Q분류 및 결과처리 단계에서는 추출된 P표본으로 하여금 Q표본을 분류하게 하고 코딩을 거쳐 PQMethod 소프트웨어를 활용하여 분석한다.

해석 및 유형화 단계에서는 분석결과를 분류하고 그 결과를 해석하여 유형화하고 유형별로 명명(naming)한다. 결론 부분에서는 FGI와 Q방법론을 활용한 연구의 결과를 종합적으로 고찰하고 시사점을 도출한다.

<표 3> Q방법론의 5단계 프로세스



IV. 결과

1. 개념정의 및 요소(element) 추출

선행연구 및 문헌분석 등 이론연구를 통하여 빅데이터 거버넌스의 개념 정의 및 7개의 구성요소를 추출하였다.

1차적으로 이론연구를 통해 빅데이터 거버넌스의 개념을 종합하여 ‘빅데이터 관련 정책·지침·표준·전략 및 방향을 수립하고 조직의 목표달성 및 가치창출을 위한 통합시스템’이라고 정의하였다. 2차에서는 1차에서 도출된 개념 정의를 바탕으로 최종 정의를 도출하기 위해 3년 이상의 현장실무자 8명을 대상으로 2021년 10월 22~23일 2일간 FGI를 실시하였다. 2차에 걸친 이론연구 및 인터뷰를 통해 빅데이터 거버넌스를 ‘빅데이터 관련 전략 수립과 목표달성 및 가치창출을 위한 협업 및 통합관리체제’로 최종 정의하였다.

<표 4> 빅데이터 거버넌스의 개념 정의

단계	목적	조사방법	도출 결과
1차	개념 정의	<ul style="list-style-type: none"> 선행연구 문헌분석 	빅데이터 관련 정책·지침·표준·전략·방향을 수립하고 조직의 목표달성 및 가치창출을 위한 통합시스템
2차	개념 재정의	FGI	빅데이터 관련 전략 수립과 목표달성 및 가치창출을 위한 협업 및 통합관리체제

현장에서의 경험과 이론을 토대로 정의된 빅데이터 거버넌스에 대한 개념 정의를 바탕으로, 구성개념 도출을 위한 요소를 추출하였다. 1차적으로 선행연구, 문헌분석 및 국내외의 4개의 데이터 거버넌스

모델(한국정보화진흥원의 Big-CAT, IDC의 클라우드 성숙도모형, Oracle의 IMRA모형, DAM 성숙도모형) 등의 사례조사를 통해 24개의 요소를 추출하였다. 1차적으로 추출된 결과를 기반으로 앞서 개념 정의에 참여했던 현장실무자 8명을 대상으로 2021년 10월 25~26일 2일간 FGI를 실시하였다. 중복되거나 애매한 표현은 통합, 삭제한 뒤 사전인터뷰를 포함하여 2차에 걸친 반구조화된 인터뷰형식을 수행하였다. 분석결과 12개의 요소로 병합, 압축되었으며 최종적으로 7개의 요소가 추출되었다. 최종 구성된 요소는 <표 5>와 같다.

<표 5> RGT를 위한 요소 추출

단계	조사방법	추출 요소
1차	<ul style="list-style-type: none"> 선행연구 문헌분석 사례조사 	전략수집, 전략추진, 데이터 분석, 조직, 인적자원, 프로세스, 인프라, 모니터링, 목표수립, 조직문화, 디지털리더십, 디지털리더십, 품질관리, R&R, 평가지표, 정책, 시스템, 도구, 보안, 윤리, 표준화, CDO, 인력배치, 육성
2차	FGI	HR(인력배치/육성/평가) 정책(전략, 목표수립) 데이터 관리(보안, 시스템) 분석기술(분석방법, 분석도구) 모니터링(평가, 보고) 조직(R&R, 문화, 리더십) 프로세스(체계, 규정)

2. RGT에 의한 구성개념(construct) 도출

이론연구 및 FGI를 통해 최종적으로 3개의 요소를 추출 후, 빅데이터 분야 5년 이상의 현장실무자 8명을 대상으로 RGT를 실시하였다.

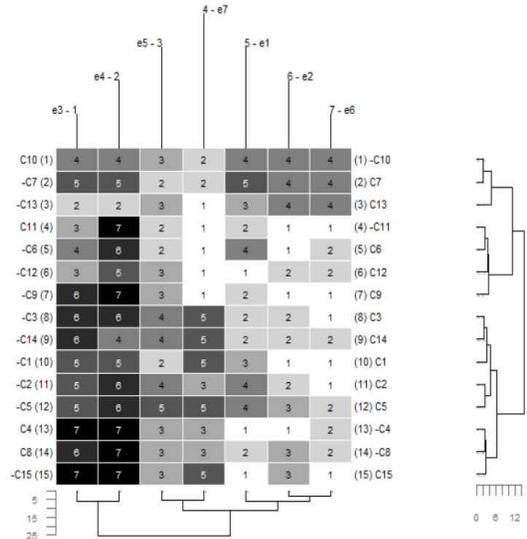
빅데이터 거버넌스에 대한 선행연구가 거의 이루어지지 않았고, 활용 분야 및 표준별로 빅데이터 관련 용어가 약간씩 다르게 쓰이는 경우가 있어 요소별로 포괄된 유사요소 혹은 하위요소들을 함께 제시하였다. 도출된 구성개념은 총 18개였으며, 중복 및 불명확한 개념을 제거, 통합하여 최종적으로 15개의 구성개념이 도출되었다. 최종 추출된 구성개념은 <표 6>과 같다.

<표 6> 구성개념(construct) 도출

구성개념	대립개념
커뮤니케이션이 중요함	C1 커뮤니케이션 비중이 낮음
co-work이 중요함	C2 개인 역량이 중요함
IT기술력	C3 통찰력 및 휴리스틱
...	...
조직구성원에 대한 관리	C14 업무에 대한 감독, 관리
인력 관리 및 통제	C15 데이터 관리 및 통제

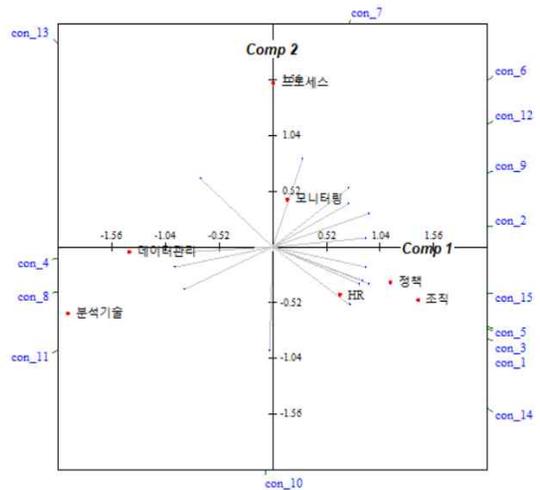
구성개념 추출이 완료된 후, 참여자들에게 6개의 요소를 전체 15개의 구성개념 관점에서 7점 척도로 양극성 평가(bipolar rating)를 실시하였다. 평정 결과는 분석프로그램에 적합한 형태로 코딩(*.xlsx, *.grd)하여 정량화하였다. R 프로그램 및 Idiogrid, OpenRepGrid를 사용하여 요인분석, 군집분석 등을 실시하였으며, 분석 결과를 바탕으로 빅데이터 거버넌스를 구성하는 주요요인을 재정렬하는 과정을 수행하였다.

먼저, OpenRepGrid를 사용하여 평정결과를 분석하였다. 전 단계에서 추출된 요소 및 RGT를 통해 도출된 구성개념을 Dendrogram으로 시각화하였고, 계층적 군집분석을 통해 구성개념의 재정렬 및 군집에 따른 대략적인 계층구조를 확인하였다. 요소는 2~4개의 군집으로 분류되었으며, 구성개념은 2~10개의 군집으로 분류 가능하였다.



<그림 2> Bertin display analysis 및 Dendrogram에 의한 Hierarchical Clustering

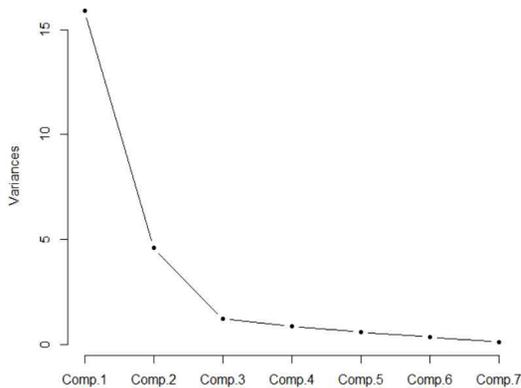
RGT의 가장 큰 장점 중의 하나는 요소와 구성개념을 동시에 공간적으로 표현할 수 있다는 것이며, <그림 3>과 같이 요소별 구성개념(construct) 항목을 Biplot analysis로 시각화하여 평정 패턴의 유사성을 파악하였다.



<그림 3> Biplot analysis result

행과 열에 따라 평정된 패턴이 특정한 패턴을 지닐 때 이를 ‘성분(component)’이라는 개념으로 정의한다. 방대한 데이터를 유사한 패턴끼리 성분으로 구분하여 분석하기 쉬운 형태로 축소시킨다. 이를 통해 RGT 자료의 구조이자 참여자의 구성체계의 구조를 확인할 수 있다(김소희, 유금란, 2018). 구성 개념 평정 패턴은 직선으로 제시되는데, 이들 직선간 각도가 작을수록 평정이 유사하고, 각도가 클수록 평정에 차이가 있음을 의미한다.

2차원의 공간 안에 제1성분에 가장 적합한 직선이 나타나고, 두 번째로 적합한 제2성분이 제1성분에 직각으로 나타난다. 제1성분에 가장 많은 영향을 주고 있는 요소는 ‘데이터 관리’, ‘분석기술’로 나타났다. 구성개념은 ‘IT기술력’, ‘IT프로세스’, ‘빅데이터분석 기술’로 나타났다. 제2성분에 가장 많은 영향을 미친 요소는 ‘프로세스’, ‘모니터링’이었고, 구성개념은 ‘가변성 반영’, ‘프로세스 전반적 관여’로 나타났다. 제1성분 설명량은 46.1%, 제2성분 설명량은 37.4%로 분석되었다.



<그림 4> Scree plot

Scree plot으로 주성분의 개수를 파악하였다. <그림 4>에서처럼 구성개념에 의한 주성분 요인은 3개로 분류할 수 있으며, 주성분을 추정하기 위해

다음과 같이 요인분석을 실시하였다.

<표 7> 요인분석에 의한 결과 요약

	fa1	fa2	fa3
SS loadings	2.188	1.892	1.819
Proportion Var.	0.313	0.270	0.260
Cumulative Var.	0.313	0.583	0.843

	fa1	fa2	fa3	Uniq.
e1	.707	-.342	.243	.325
e2	.752	-.415	.394	.107
e3	-.308	.771	.162	.284
e4	-.473	.813	.332	.005
e5	.400	-.176	.663	.370
e6	.786	-.558	.258	.005
e7	.165		.981	.005

요인구조의 정확성보다는 변수들의 군집성을 발견하기 위해 Varimax 방식으로 요인회전을 실시하였다. 분석한 결과는 <표 7>과 같다. 요인1이 총 분산의 31.3%, 요인2가 27.0%, 요인3이 26.0%를 설명하여 세 요인에 의해 설명되는 변동은 총 변동의 84.3%이다. 요인1의 경우 ‘HR(e1)’, ‘정책(e2)’, ‘조직(e6)’, 요인2의 경우 ‘데이터 관리(e3)’, ‘분석기술(e4)’, 요인3은 ‘모니터링(e5)’, ‘프로세스(e7)’의 비중이 큰 요인으로 나타났다. RGT 평정 결과를 분석한 Hierarchical Clustering, Biplot, Dendrogram, Scree plot 및 요인분석 결과를 종합하였을 때, 요인1은 ‘조직 및 HR’, 요인2는 ‘데이터 분석 및 관리’, 요인3은 ‘프로세스 및 성과관리’에 해당하는 것으로 해석해볼 수 있다.

본 RGT 평정결과 분석을 통해 빅데이터 거버넌스의 구성요소 및 구성개념을 재정렬 및 분류하였으며, 이를 기반으로 Q방법론에 의한 연구를 수행

하였다.

3. Q방법론에 의한 주요 구성요소 도출

RGT를 통해 도출된 구성개념을 바탕으로 최종 47개의 Q표본을 추출하였고, 10인의 P표본이 Q분류한 결과를 정량화하여 PQMethod로 분석하였다. 유형 추출방법으로는 주성분분석 방법을 활용하였으며, 아이겐값(Eigen Value) 1이상을 기준으로 유형의 수를 판단하였다. 마지막으로, Q진술문의 표준점수의 최고점과 최저점을 기준으로 구분된 유형의 특성을 파악할 수 있었다.

유형의 개수를 판단하기 위해 아이겐값을 확인한 결과, 아래의 <표 8>과 같이 총 세 개의 유형으로 도출되었다. 각 유형의 아이겐값은 각각 1유형이 3.995, 2유형이 2.446, 3유형이 1.136이며, 설명력은 75%로 나타났다.

<표 8> P표본의 아이겐값(eigenvalues)

Type	I	II	III
Eigenvalues	3.995	2.446	1.135
% expl.Var	40%	24%	11%
cum% expl.Var	40%	64%	75%

각 유형의 상관관계는 각 유형간의 유사성 정도를 보여주는 것으로 제1유형과 제2유형은 -.2438, 제1유형과 제3유형은 -.6018, 제2유형과 제3유형은 .5761로 나타났으며 <표 9>와 같다. R방법론에서는 요인간의 상관계수와 상관관계가 비례하고, 계수가 0에 가까울수록 독립적인 관계임을 나타내지만, Q에서는 높은 상관계수가 반드시 독립적이지 않다는 의미를 나타내는 것은 아니며, 오히려 이론적 개념들의 해석을 통하여 유형을 분류할 수 있으므로 해석의 여지는 남아있다(김홍규, 2008).

<표 9> 유형 간 상관계수

Type	I	II	III
I	1.000	-.2438	-.6018
II		1.000	.5761
III			1.000

P표본들은 다음 <표 10>와 같이 세 가지 유형으로 구분되었다. 요인가중치가 높을수록 그 유형을 대표하는 P표본으로 나타난다. 전체 P표본 10명 중 제1유형은 총 4명(40%)으로 P2가 .8795로 가장 높게 나타났으며, 제2유형은 총 4명(40%)으로 P9가 .9264로 가장 높게 나타났고, 제3유형은 총 2명(20%)이며 P4가 .8237로 가장 높게 나타났다.

요인1은 변이의 29%, 요인2는 26%, 요인3은 15%를 설명하고 있어 3개의 Q요인은 종합적으로 변이의 70%를 설명하는 것으로 분석되었다.

<표 10> Q표본의 요인분석표(factor loading)

P sample		Type 1	Type 2	Type 3
P1	IT전략실장 여	.8637x	-.2792	-.3264
P2	CDO 남	.8795x	-.1804	-.3075
P3	데이터분석팀장 남	-.3129	.3450	.6030x
P4	데이터사이언티스트 남	-.3736	.3373	.8237x
P5	정보센터장 남	.2255	.5550x	.2576
P6	책임연구원 여	-.0490	.8138x	.2752
P7	기획처장 남	.7473x	.2139	-.1940
P8	경영지원실장 남	.6845x	.3676	.0558
P9	빅데이터과교수 남	-.1941	.9264x	.2393
P10	데이터정책박사 여	.1091	.5901x	.0806
% expl.Var		29%	26%	15%

* 각 Q요인과 상관관계가 높은 참여자는 x표시

자료분석의 결과 Q유형은 총 3개로 나타났다. 해당 유형의 특성을 파악하기 위해 해당 진술문에 대

한 표준점수(z-score)를 확인하였다. 표준점수가 높은 항목(z-score > +1)을 추출해서 각 유형의 특성을 도출하였다. 본 연구는 유형별 선호도를 평가하는 것이 아니라, 인식의 차이에 따른 유형별 중요도를 파악하기 위한 것이므로, 표준점수가 낮은 항목은 무의미하다.

유형1은 빅데이터 관련 직무 경력 8년 이상의 4명의 참여자로 구성되었으며, 이들은 IT전략실장, CDO(Chief Data Officer, 최고데이터책임자), 데이터분석기획 차장, IT 경영지원실장 4명이었다. 유형1에 해당하는 진술문의 분석 결과는 <표 11>과 같다.

<표 11> 유형1의 진술문 분석 결과

유형1: 정책 및 조직관리			
진술문	해당요소	구성개념	Z-score
Q40	정책	전략	1.794
Q41	정책	전략	1.567
Q10	조직관리	관리/통제	1.467
Q20	정책	기획	1.426
Q32	조직관리	R&R	1.368
Q31	조직관리	조직문화	1.278
Q27	조직관리	관리/통제	1.264

첫 번째로 나타나는 유형1은 ‘정책 및 조직관리 중시형’으로 명명하였다. 유형1을 특징짓는 진술문들이 주로 기획, 전략으로 구성된 ‘정책’요소와 조직문화, R&R, 관리 및 통제로 구성된 ‘조직관리’와 연관되어있기 때문이다. 또한 각각의 유형 내에서 P표본의 요인가중치가 가장 높은 P표본일수록 해당 유형을 가장 잘 설명할 수 있는 전형 혹은 대표적인 사람임을 나타내는데, 기업의 CDO가 가장 높은 점수를 부여한 진술문 Q31(빅데이터 관련 부서간 수평적인 조직문화 조성이 수반되어야 한다), Q40(우리 조직의 비전 및 가치에 최적화된 빅데이터

거버넌스의 모델이 수립되어야 한다), Q20(비즈니스 요구에 대해 빅데이터를 분석하여 기획, 관리하는 것이 중요하다) 3가지와도 일치함을 확인할 수 있었다. 유형1에 해당하는 P표본들은 데이터 관련 기획 및 전략, 조직을 경영하는 직무를 담당하고 있어 빅데이터 거버넌스에 정책 및 조직관리가 주요 요소로 구성되어야 함을 알 수 있다.

<표 12> 유형2의 진술문 분석 결과

유형2: 성과지향 프로세스			
진술문	해당요소	구성개념	Z-score
Q30	프로세스	진단/개선	1.610
Q07	프로세스	기술적 체계	1.587
Q38	모니터링	성과관리	1.462
Q28	모니터링	성과평가	1.452
Q25	조직관리	R&R	1.230
Q47	프로세스	진단/개선	1.132
Q34	HR	리더십	1.107

유형2는 빅데이터 관련 직무 경력 12년 이상의 4명의 참여자로 구성되었으며, 이들은 정보센터장, 빅데이터연구소 책임연구원, 빅데이터과 교수, 데이터정책전공 박사 등 4명이었다. 유형2에 해당하는 진술문의 분석 결과는 <표 12>와 같다.

유형2는 ‘성과지향 프로세스 중시형’으로 명명하였다. 유형2를 대표하는 진술문들이 주로 진단 및 개선, 기술적 체계로 구성된 ‘프로세스’ 요소와 성과관리, 성과평가 등으로 구성된 ‘모니터링’과 연관되기 때문이다. 해당 유형의 요인가중치가 가장 높은 빅데이터과 교수가 가장 높은 점수를 부여한 진술문 Q30(빅데이터 분석 수행을 위한 현 수준 진단 및 개선점 파악 등 분석 준비도를 진단하는 것이 중요하다), Q47(각 프로세스의 진단 및 개선이 지속적으로 이루어져야 한다)의 2가지와도 일치함을 확인할 수 있었다. 유형2에 해당하는 P표본들은 협력

부서 간 업무체계, 분석 준비도 및 성숙도 진단, 빅데이터 분석 프로세스 모니터링을 담당하고 있어 빅데이터 거버넌스에 프로세스, 모니터링이 주요요소로 구성되어야 함을 알 수 있다.

<표 13> 유형3의 진술문 분석 결과

유형3: 데이터 분석 및 관리			
진술문	해당요소	구성개념	Z-score
Q06	데이터관리	데이터모니터링	1.704
Q13	데이터관리	데이터모니터링	1.581
Q21	데이터관리	데이터분류	1.581
Q37	데이터분석	분석기술	1.491
Q46	데이터분석	분석인프라	1.371
Q33	조직관리	조직구성	1.371
Q14	조직관리	인력배치	1.371

Flagged with * are all Non-significant at P>0.5

유형3은 빅데이터 관련 직무 경력 7년 이상의 2명의 참여자로 구성되었으며, 이들은 데이터분석팀장, 데이터 사이언티스트 등 2명이었다. 유형3에 해당하는 진술문의 분석 결과는 <표 13>과 같다.

유형3은 ‘데이터 분석 및 관리’로 명명하였다. 유형3을 대표하는 진술문들이 주로 데이터 모니터링 및 데이터 분류로 구성된 ‘데이터 관리’ 요소와 분석기술, 분석인프라 등으로 구성된 ‘데이터 분석’과 연관되기 때문이다. 또한 추가적으로 도출된 ‘조직 관리’는 실제 빅데이터를 분석하는 데이터사이언티스트와 데이터 분석 조직의 총괄담당자가 실제 데이터 분석과 관리를 하는 전문 인력에 대한 중요도와 요구도가 높음을 보여준다.

해당 유형의 요인가중치가 가장 높은 데이터사이언티스트가 가장 높은 점수를 부여한 진술문 Q6(수행과정에서 분석자료의 가변성을 반영해야 한다), Q13(데이터에 대한 실시간 관리 및 통제가 중요하다), Q37(최신 분석기술 및 도구에 대한 지원이 원

활히 이루어져야 한다)의 3가지와도 일치함을 확인할 수 있었다. 유형3에 해당하는 P표본들은 빅데이터 분석, 데이터 통합관리 등을 담당하고 있어 빅데이터 거버넌스에 데이터 관리 및 분석이 구성되어야 함을 알 수 있다.

<표 14> 공통(Consensus) 요인

진술문	해당요소	구성개념	요인가		
			유형1	유형2	유형3
Q26*	모니터링	지속적 환류	2	2	2
Q32*	조직관리	R&R	2	2	1
Q17	HR	IT리더십	0	2	2

유형1~3은 각 진술문에 대한 인식의 차이를 통해 유형별로 구분하여 그 특징을 논할 수 있었다. Q방법론에서는 유형 간 차이를 보여주기도 하지만 유사하게 인식하는 주요 공통요인 역시 확인할 수 있다는 장점이 있다. 본 연구 또한 유형 간 유사한 의견을 보인 공통요인도 확인할 수 있었는데, 해당 내용은 <표 14>와 같다.

진술문 Q26의 ‘지속적인 모니터링을 통한 환류시스템이 원활히 이루어져야 한다’, Q32의 ‘조직 및 각 부서 간 R&R이 명확해야 한다’, Q17의 ‘IT리더십이 중요하다’의 3개의 진술문이 모든 유형에서 중요한 구성요소로 인식되고 있음을 알 수 있다. 빅데이터 거버넌스의 구축 및 완성도를 향상시키기 위한 구성요소로 거버넌스의 인적, 물적 인프라 및 시스템에 대한 전반적인 모니터링 및 지속적인 환류가 중요하며, 거버넌스를 구성하는 각 조직 간 갈등 및 책임소재 불명에 대한 소지를 줄이기 위한 R&R의 명확성이 필요함을 파악할 수 있다. 특히 진술문 Q17의 ‘IT리더십’은 의사결정권자들의 기본적인 IT 및 데이터분석 관련 기본 역량이 빅데이터 거버넌스의 궁극적인 목표인 조직의 목표달성 및 가치창출 가능케 할 수 있는 중요한 요소임을 증명해주고 있다.

4. 최종 주요 구성요소 도출

각 단계의 도출 결과를 종합, 분석하여 최종적으로 빅데이터 거버넌스의 주요 구성요소를 개발하였으며, 결과는 <표 15>로 정리하였다.

<표 15> 빅데이터 거버넌스의 최종 주요 구성요소

연구방법	도출내용	도출 결과
<ul style="list-style-type: none"> • 선행연구 • 문헌분석 • 사례조사 • FGI 	<ul style="list-style-type: none"> • 개념 정의 • 요소 (elements) 	<ul style="list-style-type: none"> • 빅데이터 관련 전략 수립과 목표달성 및 가치 창출을 위한 협업 및 통합관리체제 • HR(인력배치/육성/평가), 정책(전략, 목표수립), 데이터 관리(보안, 시스템), 분석기술(분석방법, 분석도구), 모니터링(평가, 보고), 조직(R&R, 문화, 리더십), 프로세스(체계, 규정)
		▼
RGT	<ul style="list-style-type: none"> • 구성개념 (constructs) 	<ul style="list-style-type: none"> • 정책, 조직관리, HR, 프로세스, 모니터링, 데이터관리, 데이터 분석
		▼
Q 방법론	<ul style="list-style-type: none"> • 하위요소 • 인식유형 	<ul style="list-style-type: none"> • R&R, 조직구성, 전략, 기획, 조직문화, 기술적 체계, 인력 배치, 진단 및 개선, 데이터 분류, 성과평가, 데이터모니터링, IT리더십, 분석기술, 분석인프라, 지속적 환류 • 정책 및 조직관리 중시형, 성과지향 프로세스 중시형, 데이터분석 및 관리 중시형
		▼
최종 분석	<ul style="list-style-type: none"> • 주요 구성요소 	<ul style="list-style-type: none"> • 정책 • 조직관리 • 프로세스 • 데이터 분석 및 관리

IV. 논의

빅데이터 분석이 일회성 단기 프로젝트로 끝나지 않고, 환류와 개선을 통한 업무 최적화와 지속적인 혁신과 성장을 실현해 나가기 위해서는 협력분야별 거버넌스 체계가 필요하다. 즉, 빅데이터 분석과 활용을 위한 다양한 모델의 도입과 정책 과정에서 생성되는 여러 변화요인(갈등과 저항, 책임 등)을 미리 파악하여 지원할 수 있어야 하며, 이를 위해서는 빅데이터 거버넌스의 구체적인 개념 정의와 주요 구성 요소 정립이 선행되어야 한다. 본 연구는 이를 위해 데이터 거버넌스에 관한 선행연구 및 문헌들을 분석하고, 분석결과를 기반으로 RGT기법 및 Q방법론을 통해 빅데이터 거버넌스의 개념 정의와 구성요소를 도출할 수 있었다. 연구결과를 요약하면 다음과 같다.

첫째, 빅데이터 거버넌스의 개념은 ‘빅데이터 관련 전략 수립과 목표달성 및 가치 창출을 위한 협업 및 통합관리체제’로 정의하였다.

둘째, 빅데이터 거버넌스의 구성요소로는 정책, 조직관리, 프로세스, 데이터 분석, 데이터 관리, 모니터링 등으로 나타났다. 이들의 하위요소로는 전략, 기획, R&R, 진단/개선, 분석 인프라, 분석 기술 지원, 조직구성, 인력배치, 지속적 환류, 리더십 등으로 구성되었다.

셋째, 빅데이터 거버넌스를 구성하는 조직 및 구성원들은 ‘정책 및 조직관리 중시형’, ‘성과지향 프로세스 중시형’, ‘데이터 분석 및 관리 중시형’의 세 가지 집단으로 구분되었다.

넷째, 도출된 구성요소와 유형분류를 종합하여 ‘정책’, ‘조직관리’, ‘프로세스’, ‘데이터 분석 및 관리’의 4가지 주요 구성요소로 최종 도출하였다.

연구 결과를 통한 논의 및 시사점은 다음과 같다. 첫째, 빅데이터 거버넌스의 주요 구성요소 및 하위 요소는 선행연구 및 RGT기법, Q방법론 연구 간 많은 유사성을 보였지만, 선행연구에서 도출하지 못한

중요한 구성요소로 ‘IT리더십’에 주목해 볼 필요가 있다. 이는 빅데이터 활용도 및 거버넌스 성숙도가 높은 국외 기업들에서 기존의 ‘CIO(Chief Information Officer)’의 역할에 데이터 관련 역량을 보완하여 고도화된 ‘CDO(Chief Data Officer)’의 직급이 급격히 증가한 것과 유사한 관점에서 생각해 볼 수 있다. 미국 NewVantage Partners(2021)에 따르면, Fortune 1000대 기업의 68%에서 이미 CDO를 임명하였으며, 이들의 주요 직무는 혁신, 데이터 거버넌스, 애널리틱스였다. 4차 산업혁명의 핵심은 역동적으로 변화하는 환경에서 빅데이터를 통해 가치있는 결과를 추출하여 얻어낸 인사이트로 변화에 대한 적절한 대응을 하는 것이다. 이를 위해 각 기업 및 기관에서는 데이터에 대한 이해와 관리, 비즈니스를 위한 전략적 자산으로 변환, 활용하는 능력이 매우 중요한데, 이를 전담하는 직책이 CDO인 것이다. 또한 ‘C’ 레벨이 아니더라도 최소한의 의사결정권을 소유한 직책의 담당자들은 IT 및 데이터와 관련된 기본적인 문해력이 요구되고 있다. 이번 연구를 통해 국내에서도 빅데이터 거버넌스의 구축을 위한 매우 중요한 기본 요소로 인식되고 있음을 알 수 있다.

둘째, 기존의 선행연구에서 주로 실시되었던 질적, 양적으로 양분된 연구방법론에서 벗어나, 이론 및 현장연구로 구분하고, 현장연구는 다시 관련 실무자 단계와 전문가 단계로 구분하여 심층인터뷰 및 양적 분석 방법을 동시에 적용하였다는데 연구의 학문적 기여를 찾을 수 있다. 국내외 선행연구 및 문헌자료를 분석하여 기초자료를 추출하고, 이를 기반으로 현장 실무자들을 대상으로 심층인터뷰 및 RGT기법을 통한 인지구조 분석으로 빅데이터 거버넌스의 정의 및 요소와 구성개념을 도출하였다. 마지막 단계로 이를 Q연구방법론의 기초자료로 활용하여 빅데이터 관련 전문가 단계에서 인식 및 통찰을 통한 유형화를 거쳐 최종 빅데이터 거버넌스의 주요 구성요소를 검증, 도출할 수 있었다. 이는 질적 및 양적이 혼합된 연구방법을 적용하여 이론연

구 뿐만 아니라 현장실무자 및 전문가 대상의 실증 연구에 의한 결론을 이끌어 냈다는데서 학문적, 실용적 의의를 찾을 수 있다.

단, 본 연구는 국내에서 거의 수행되지 않은 빅데이터 거버넌스 관련 연구로 인해 이론연구는 상당부분이 국외의 사례였다는 점과 국내의 경우에는 IT거버넌스 및 데이터 거버넌스 모델 연구에서 기초자료를 추출했다는 점을 한계점으로 들 수 있다. 하지만 이러한 한계점에도 불구하고, 여전히 초기단계에 불과하지만 국내 빅데이터 생태계 구축 시도에 본 연구결과가 기초자료로 활용가능할 것이며, 향후 더 많은 연구 참여자들이 보완되어 빅데이터 거버넌스의 고도화된 연구를 수행할 수 있을 것이다. 또한 이를 활용하여 빅데이터 거버넌스의 정성적, 정량적 진단이 가능한 모델 수립을 위한 연구가 수행될 것으로 기대한다.

참고문헌

- 길병욱, 이소희, 이송이, 정희정 (2020). Q방법론의 이해와 적용. 대전: 충남대학교출판문화원.
- 김상배 (2015). 빅데이터의 국가전략; 21세기 신흥 권력 경쟁의 개념적 성찰. 국가전략, 21(3), 5-36.
- 김석수 (2011). EA를 위한 데이터 거버넌스 구축 사례 연구. 한국엔터프라이즈아키텍처학회, 8(3), 255-265.
- 김소희, 유금란. (2018). 심리학 연구에서 개인구성 심리학 연구법의 적용: 레퍼토리 그리드 기법을 중심으로. 한국심리학회지: 37(3), 349-383.
- 김순은 (2010). Q 방법론의 이론과 철학. 한국사회와행정연구, 20(4), 1-25.
- 김홍규 (199 6). Q 방법론의 유용성 연구. 주관성 연구, 1, 15-33.
- 박정, 조완섭 (2020). 교육중단연구 분석을 위한 빅데이터 플랫폼 개발 및 적용. 한국빅데이터학회,

- 5(1), 11-27.
- 백평구 (2015). Q모집단으로부터 Q표본 구성 과정에 대한 분석: 학술지 주관성 연구 게재 논문을 중심으로(1996~2014). *주관성연구*, 30, 109-130.
- 송경재, 장우영, 조인호 (2018). 빅데이터 거버넌스의 가능성과 과제에 관한 탐색. *사회이론*, 53, 153-187.
- 이정규, 정승렬. (2019). 프로세스 거버넌스 메타모델과 프레임워크. *인터넷정보학회논문지*, 20(4), 63-72.
- 장경애 (2017). 고객의 요구사항에 기반한 데이터 품질 평가속성 및 우선순위 도출. *정보처리학회*, 4(12), 549-560.
- 정부부처합동 (2021). 한국판 뉴딜 2.0. 정부정책자료(2021.07.14.).
- 조완섭 (2017). 빅데이터 거버넌스와 표준화 동향. *개방형컴퓨터통신연구회*, 30(2), 26-29.
- 최명석, 이상환 (2020). 데이터 관리 계획의 국내 현황 및 과제. *한국콘텐츠학회*, 20(6), 220-229.
- 최영환, 조완섭, 이경희 (2018). 스마트 물관리를 위한 빅데이터 거버넌스 모델. *한국빅데이터학회*, 3(2), 1-10.
- 최창호, 유연우 (2017). 탐색적 요인분석과 확인적 요인분석의 비교에 관한 연구. *디지털정책학회*, 15(10), 103-111.
- 현영란 (2015). 지방정부역량이 지방정부와 비영리 조직간 협력적 계약관계에 미치는 영향. *한국거버넌스학회보*, 22(2), 259-281.
- 한국데이터진흥원 (2019). 데이터 분석 전문가 가이드(ADP). 개정판 6쇄, 서울: 한국데이터베이스진흥원.
- Akoka, J., Comyn-Wattiau, I., & Laoufi, N. (2017). Research on Big Data - A systematic mapping study. *Computer Standards & Interfaces*, 54, 105-115.
- Al-Badi, A., Tarhini, A., & Khan, A. I. (2018). Exploring big data governance frameworks. *Procedia computer science*, 141, 271-277.
- Bloomberg, L. D., & Volpe, M. (2008). *Completing your qualitative dissertation: A roadmap from beginning to end*. Los Angeles: Sage.
- Chandler, N., Hostmann, B., Rayner, N., & Herschel, G. (2011). *Gartner's Business Analytics Framework*, Gartner. Inc., Stanford.
- Daniel, B. K., & Butson, R. (2013). Technology Enhanced Analytics (TEA) in Higher Education, *Proceedings of the International Conference on Educational Technologies*, 29 November-December, 2013, Kuala Lumpur, Malaysia, 89-96.
- Daniel, B. K. (2019). Big Data and data science: A critical review of issues for educational research. *British Journal of Educational Technology*, 50(1), 101-113.
- Edjlali, R., & Friedman, T. (2011). *Data quality for big data: Principles remain, but tactics change*. Gartner.
- Fransella, F., Bell, R., & Bannister, D. (2004). *A manual for repertory grid technique*. John Wiley & Sons.
- Heeks, R. (2001). *Reinventing government in the information age: International practice in IT-enabled public sector reform*. Vol. 1. London: Psychology Press.
- Jagadish, H. V., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J. M., Ramakrishnan, R., & Shahabi, C. (2014). Big data and its technical challenges. *Communications of the ACM*, 57(7), 86-94.
- Kelly, G. (1955). *The psychology of personal constructs*. England: Oxford.

- Kelle, O. (2015). Top 10 artifacts needed for data governance. *First San Francisco Partners*.
- Kim, H. Y., & Cho, J. S. (2018). Data governance framework for big data implementation with NPS Case Analysis in Korea. *Journal of Business and Retail Management Research*, 12(3).
- Kitchin, R. (2014). *The data revolution. Big data, open data, data infrastructures & their consequences*. London: Sage.
- Ladley, J. (2012). *Data governance*. Massachusetts: Morgan Kaufmann.
- Laney, D. (2001), “3-D Data Management: Controlling Data Volume, Velocity and Variety,” META Group Research Note, February 6.
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing research*, 35.
- NewVantage Partners (2021). *The Journey to Becoming Data-Driven: A Progress Report in the State of Corporate Data Initiatives*. Big Data and AI Executive Survey 2021. New Vantage Partners LLC. 1-16.
- Oracle(2021). *What is Bigdata?* retrieved from <https://www.oracle.com/big-data/what-is-big-data>.
- Roberts, L. D., Chang, V., Gibson, D. (2017). Ethical Considerations in Adopting a University- and System-Wide Approach to Data and Learning Analytics. In Kei Daniel B. (Eds.), *Big Data and Learning Analytics in Higher Education*(89-108). Switzerland: Springer.
- Romero, S. (2013). *IT 커버넌스*(서기운 역). 서울: 한빛미디어.(원전 2011에 출판)
- Soares, S. (2012). *Big Data Governance*. Chicago: Chicago Review Press.
- Soares, S. (2013). *IBM InfoSphere: A platform for Big Data governance and process data governance*. Mc Press.
- Soares, S. (2018). *The IBM Data Governance Unified Process*. MC Press.
- SO/IEC38500 (2015). *Information Technology—Governance of IT-For the organization*.
- Stenner, P. (2009). *What is a Q Methodology?*. University of Brighton.
- Stephenson, W. (1953). *The study of behavior: Q-technique and its methodology*. Chicago: University of Chicago Press.
- Violino, B. (2021., JUL 8). *6 best practices for good data governance*. Retrieved from <https://www.cio.com/article/3624648/6-best-practices-for-good-data-governance.html>
- 투고일자 : 2021. 11. 14.
 심사일자 : 2021. 12. 14.
 게재확정일자 : 2021. 12. 31.

Development of Major Components of Big Data Governance Using RGT and Q methodology

JiYeon Kang

Hanyang University

The purpose of this study is to define the concept and develop major components of big data governance to establish efficient and effective big data governance. To this end, basic data were collected through prior research, literature, and case analysis by dividing theoretical and field studies, and based on this, RGT techniques and Q methodology mixed with qualitative and quantitative research methods were sequentially applied to derive research results. As a result of the study, first, big data governance could be defined as a 'collaboration and integrated management system to establish strategies related to big data, achieve goals, and create values'. Second, the main components were 'policy', 'organizational management', 'process', and 'data analysis and management'. Third, there were three types of recognition of big data governance: 'policy and organizational management-centric type', 'performance-oriented process-centric type', and 'data analysis and management-centric type'. In this study, theoretical and empirical studies were conducted sequentially, and the academic contribution of the study can be found in that both qualitative and quantitative research methods were applied at the same time. In addition, practical significance can be found in that the results of this study can be used as basic data for the establishment of big data governance and research on the big data ecosystem.

Keywords: big data, big data governance, RGT, Q methodology