

# 국내 마이데이터 태동과 데이터 거래에 관한 잠재적 주제 분석

조지연<sup>1</sup>, 이봉규<sup>2\*</sup>

<sup>1</sup>연세대학교 방송통신정책연구소 전문연구원, <sup>2</sup>연세대학교 정보대학원 교수

## Study on Potential Topics of the MyData and Data Transactions Using LDA Topic Modeling

Ji Yeon Cho<sup>1</sup>, Bong Gyou Lee<sup>2\*</sup>

<sup>1</sup>Research Fellow, Communications Policy Research Center, Yonsei University

<sup>2</sup>Professor, Graduate School of Information, Yonsei University

요 약 최근 마이데이터 사업이 본격화됨에 따라 개인데이터 활용에 관한 관심은 더욱 증가하고 있다. 그러나 마이데이터에 관한 연구들은 아직까지 초기 단계로서, 법·제도적 논의에 집중되어 있으며 종합적인 시각에서 마이데이터를 조망한 연구는 부족한 상황이다. 이에 본 연구는 데이터거래에 관한 논의가 본격화된 2018년부터 현시점까지의 뉴스를 분석하여 사회적 논의가 형성한 잠재적 주제를 분석하고자 하였다. LDA 토픽모델링 기법을 통한 뉴스데이터 분석을 통하여 마이데이터 기반 금융혁신, 마이데이터 사업자 지위 및 허가, 데이터 관련 법 개정, 안전한 데이터 활용, 데이터 경제 활성화 정책, 사업자의 전략 등 6개의 잠재적 주제를 도출하여 제시하였다. 본 연구는 마이데이터 태동과 함께 나타난 사회적 이슈를 종합적으로 조망하고 관련 논의의 공백을 도출하여 시사점을 제시한 의의가 있다. 향후 연구에서는 마이데이터 서비스 개시 이후의 변화를 파악하고, 마이데이터 산업별 심도깊은 연구를 통하여 구체적인 시사점을 제공하길 기대한다.

주제어 : 마이데이터, 데이터거래, 이슈 분석, 토픽모델링, LDA 알고리즘

**Abstract** With the recent full-fledged MyData service, interest in the use of personal data is increasing. However, studies on MyData are still in the early stages, focusing on legal and institutional discussions, and studies from a comprehensive perspective are insufficient. Therefore, this study aimed at finding the potential topics formed by social discussions by analyzing news data from 2018 to the present. News data analysis using LDA topic modeling were conducted and 6 potential topics including digital transformation in finance, scope of Mydata business license, amendments and data-related laws, safe use of big data, data economy promotion policy and strategy of the financial industry were derived. This study has significance in that it comprehensively viewed the issues that emerged with the MyData and deriving gaps in previous discussion. Future research is expected to identify changes after the launch of MyData service and provide specific implications through research by specific industries.

**Key Words** : MyData, Data Transaction, Issue analysis, Topic Modeling, LDA algorithm

\*This work was supported(in part) by the Yonsei University Research Fund(Post Doc. Researcher Supporting Program) of 2020 (project no.: 2020-12-0068)

\*Corresponding Author : Bong Gyou Lee(bglee@yonsei.ac.kr)

Received January 9, 2022

Revised January 26, 2022

Accepted March 20, 2022

Published March 28, 2022

## 1. 서론

스마트 기술(AI, Big Data, Cloud Computing, IoT)을 토대로 한 지능정보화시대에 데이터는 가장 중요한 자원으로 주목받고 있다. 데이터를 얼마나 확보하고 잘 활용하는지는 기업의 중요한 경쟁력으로 평가되고 있으며[1], 주요국들은 해당 자원의 효율적인 배분과 활용을 위한 다양한 정책과 제도를 마련하고 있다. 우리나라의 경우 데이터를 영리 목적으로 활용하는데 법·규제의 제약으로 한계가 있다는 의견이 제기되어 왔다[2,3]. 또한, 산업 간 데이터를 공유·결합하고 거래할 수 있도록 하는 분야별 데이터거래소가 출범된 바 있으나 법·제도적 문제로 시장이 경직되어 있었고, 데이터 거래 경험에 있는 국내 기업의 비중은 28%에 불과하였다[4]. 이러한 문제의식에서 정부는 2020년 데이터 3법으로 통칭되는 개인정보보호법·정보통신망법·신용정보법을 개정하며 국내 데이터 공유 및 시장거래 활성화의 단초를 제공하였다. 또한, 정부는 ‘개인이 본인의 정보에 대한 권리를 보장받고, 정보의 주체인 개인의 동의와 의사에 따라 데이터를 제공하고 활용하는 것’을 의미하는 ‘마이데이터’ 사업을 추진하며 개인 데이터를 활용하는 생태계 활성화를 위하여 노력하고 있다[5].

이러한 마이데이터는 데이터 이용 패러다임 전환의 계기로 주목받고 있으나, 데이터 거래와 활용이 안전하고 올바르게 다루어질지에 대한 우려가 함께 제기되고 있다. 마이데이터는 2020년부터 실증 사업이 2022년 1월부터 본 서비스가 시작되었으며, 아직까지 개인이 본인의 데이터 이동에 관한 의사결정 권한을 주도적으로 행사한 적이 없는 상황이다. 사업은 아직까지 초기 단계로서 관련 논의는 법적 한계를 분석하는 연구에 집중되어 있다[6]. 점차 개인 데이터 이동과 거래가 본격화되며 데이터를 둘러싼 산업의 변화가 시작되었지만, 마이데이터에 관한 논의를 종합적인 시각에서 바라본 연구는 아직 부족한 실정이다. 이에 본 연구는 마이데이터 사업을 시작으로 데이터 거래가 본격화되는 현시점에, 관련된 전반적 논의가 어떻게 형성되어 왔는지 관련 잠재적 주제를 체계적으로 분석하는 것을 목적으로 한다. 이를 위하여 개인 데이터 활용에 관한 사회적 논의가 본격화된 2018년부터 지금까지의 뉴스데이터를 토픽모델링 기법을 통해 분석한다. 뉴스데이터는 최근 사회 현상과 이슈를 보도하며 사회적 논의 양상의 변화를 나타내는 유의미한 데이터이다[7]. 뉴스데이터 분석을 통하여 마이데이터를 중심으로 한 데이터 거래에 관한 사회적 논의의 잠

재적 주제 양상을 종합적으로 파악하고, 향후 개인의 데이터가 올바르게 활용되고 거래가 활발하게 이루어지기 위한 시사점을 제시하고자 한다.

## 2. 연구 배경

### 2.1 데이터 거래와 마이데이터 관련 연구

빅데이터 산업에 관한 관심이 고조되며 데이터 활용과 거래에 관한 연구가 지속적으로 이루어져 왔다. 기존 빅데이터 거래와 활용에 관한 연구들은 데이터 유통체계와 거래 형태, 데이터의 효율적 배분을 위한 기제를 논의한 연구[8], 데이터 시장의 구성요소에 대한 규명[3], 데이터 거래의 구성원들을 규명한 연구[9], 데이터 생태계와 데이터 거래 시장 관점에서 논의한 연구[4] 등이 있다. 이러한 기존 연구가 기업이 보유하고 있는 빅데이터 활용의 맥락에서 이루어졌다면, 데이터 3법 개정과 마이데이터 등장 이후에는 기업 간 빅데이터 거래와 개인 데이터 활용에 관한 관심이 증가하고 있다. 특히, 2020년 1월, 데이터 3법 개정안의 법제화로 개인정보 비식별화 조치가 이루어진 가명 정보는 개인의 동의 없이 상업적 목적으로 활용할 수 있게 되었다. 이를 기점으로 데이터 거래에 관한 논의가 시작되었다. 일반적으로 마이데이터란 정보주체인 개인에게 자신의 정보에 대한 권리를 보장하고, 개인의 의사결정에 따라 개인데이터를 개방하고 활용할 수 있게 하는 것을 의미한다. 가장 활발하게 추진 중인 금융 분야의 마이데이터는 고객의 동의하에 제 3자가 고객을 대신하여 여러 금융기관에 분산되어 있는 계좌 정보, 금융 거래 내역 등의 개인 금융데이터를 수집, 관리하며 지급지시 서비스를 제공하는 등의 사업을 의미한다[6]. 종합하면, 마이데이터는 정보주체인 개인의 결정에 따라 기업 간 데이터의 이동, 교환, 거래가 이루어지는 등 데이터 활용의 기반을 조성하였다고 볼 수 있다. 이러한 ‘마이데이터’를 중심으로 한 국내 연구는 아직까지 초기 단계로서 법적 제약을 논의한 연구를 중심으로 이루어져왔으며, 사용자에 관한 논의가 시작되고 있다.

데이터 3법과 함께 데이터 거래와 활용을 위한 연구는 데이터 3법 개정안 내용을 분석한 연구를 시작으로[2], 마이데이터의 법적 이슈, ‘정보이동권’과 개인의 정보 권한 등 개인의 법적 권리를 중심으로 한 연구[10,11]가 있다. 즉, 마이데이터에 관한 기존 논의는 마이데이터가 활성화되는 과정의 법·제도적 제약요인[6]과 제도적 개선점을 분석한 연구에 집중되어 있다. 또한, 일부 데이터

거래와 공유, 법, 제도, 기술적 측면의 이슈를 살펴본 연구는 금융서비스와 의료데이터 분야의 마이데이터를 중심으로 논의가 이루어졌다[13,14]. 마이데이터를 종합적인 관점에서 살펴본 연구로는 양경란 외(2021)의 연구가 있으며, 이 연구는 마이데이터에 관한 초기 연구로서 비즈니스 생태계 모델 관점에서 향후 등장할 행위자의 특징을 파악하는 연구를 진행하였다[1]. 그러나 아직 마이데이터의 태동과 함께 나타난 이슈 및 관련한 사회적 논의 주제에 관하여 종합적인 시각에서 조망한 연구는 부족한 상황이다. 따라서 본 연구는 데이터 3법 개정과 마이데이터 시행이 이루어지기까지 데이터 거래와 활용에 관한 사회적 논의가 형성한 잠재적 주제를 종합적 시각에서 살펴보고자 한다. 이를 위하여 본 연구는 마이데이터 태동과 함께 새로운 시장이 형성되는 등의 사회적 현상이 나타나고 있는 현시점에서 관련한 잠재적 주제를 토픽모델링 기법을 통한 뉴스데이터 분석을 통해 제시하고 시사점을 도출하고자 한다.

## 2.2 토픽모델링과 뉴스데이터 분석

본 연구는 빅데이터가 중요한 자본으로 주목받는 최근의 상황에서 마이데이터의 등장과 함께 개인의 데이터 활용과 거래를 둘러싼 사회적 논의가 어떠한 잠재적 주제를 형성하고 있는지 분석하고자 한다. 이러한 연구목적 달성을 위하여 텍스트 마이닝 기법을 통해 관련 뉴스데이터를 분석하고, 잠재적 주제를 도출하여 제시한다. 텍스트 마이닝 기법의 하나인 토픽모델링은 문서를 구성하고 있는 키워드를 분석하여 문서가 가지고 있는 주제(topic)를 도출하는데 이용되는 분석기법이다. 토픽모델링 기법 중 LDA(Latent Dirichlet Allocation)는 비정형 텍스트 데이터로부터 의미 있고 일관성 있는 주제를 도출해내는 기법으로 많이 적용되어 왔으며[15], 주요 문제, 요인, 연구 동향 및 주제 구조를 분류하는 가장 보편적인 접근 방식이다[16]. LDA는 문서에서 잠재적 의미를 내포하고 있는 키워드의 출현 확률을 토대로 문서 내에 형성되어 있는 잠재적 토픽을 찾아내는 알고리즘이다.

본 연구는 마이데이터 서비스에서의 데이터 거래에 관한 사회적 논의가 어떠한 잠재적 주제를 형성하고 있는지를 체계적으로 살펴보고자 LDA 기법을 통한 뉴스데이터 분석을 수행한다. 일반적으로 뉴스는 사건, 사고와 사회경제적 주요 이슈에 대한 사실을 대중에게 전달하기도 하고, 반복적인 정보의 전달은 사회적 여론형성에 영향을 미치기도 한다. 즉, 뉴스데이터를 통해 특정 주제에

대한 사실을 파악할 뿐만 아니라, 사회적 인식과 논의의 흐름을 파악할 수 있다[17]. LDA 토픽모델링을 통한 뉴스 분석으로 주제 형성 방향을 도출하거나[15], 인공지능 기술에 관한 사회적 이슈를 분석하고[18], 특정 사안에 대한 사회적 인식을 파악[19]하는 등의 연구가 이루어져 왔다. 이처럼 기존 연구에서는 텍스트 마이닝 기법을 통한 뉴스데이터 분석은 사회적 이슈와 인식을 파악하고 해당 기사들이 형성한 잠재적 주제를 파악하는데 폭넓게 적용되어 왔다. 본 연구에서도 토픽모델링 기반 뉴스데이터 분석을 통하여 마이데이터에 관한 사회적 논의가 형성한 잠재적 주제를 파악하고 시사점을 도출한다.

## 3. 연구방법

마이데이터와 관련한 잠재적 주제를 도출하기 위한 본 연구의 절차는 Fig. 1과 같다. 마이데이터와 관련한 뉴스 기사를 수집하고, 텍스트 데이터를 추출한 후, LDA 기법을 통한 토픽모델링 분석을 수행하여 최종적으로 시사점을 도출하였다. 단계별 수행 내용은 다음과 같다.

<b>Step 1.</b> <b>Data collection</b>	Search news articles from BigKinds DB Extract words from articles
<b>Step 2. Data pre-processing</b>	Clean the collected text data Extract noun and remove stopwords
<b>Step 3. Text mining</b>	Deploy LDA on data Extract topic models
<b>Step 4. Find implications</b>	Label extracted topics Interpret main topics and derive implications

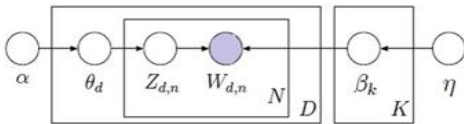
Fig. 1. Research Procedure

### 3.1 데이터 수집 및 전처리

본 연구는 한국언론진흥재단의 뉴스데이터베이스인 빅카인즈를 통해 데이터를 수집하였다. 1차적으로 뉴스 수집 범위를 전체 기간으로 ‘마이데이터’ AND ‘데이터 거래’를 검색어로 중앙지, 경제지, 방송사, 전문지에 보도된 뉴스를 수집하였다. 마이데이터와 데이터 거래의 내용을 담은 뉴스가 데이터 3법 개정 논의가 이루어진 2018년부터 본격화되었음을 확인한 후, 검색 기간을 2018년 1월 1일부터 데이터 수집 시점인 2021년 10월 30일로 한정하였으며, 총 1035건의 뉴스를 수집하였다. 최종 분석에는 중복뉴스 35건과 연구주제와 관련성이 적은 홍보성 보도, 정보 공지 뉴스 등을 제외한 662건을 포함하였다.

### 3.2 분석 방법 및 절차

뉴스데이터 분석은 Python을 통한 LDA 알고리즘을 통하여 토픽모델링 분석을 수행하였다. 우선 비정형 텍스트 데이터인 뉴스를 전처리 과정을 통하여 분석을 위한 표준화된 데이터로 정제하였다. 특정 서비스나 기관명 등의 의미 없이 반복적으로 나타나는 단어들의 불용어 처리(stop word)를 하였으며, 형태소 분석을 통하여 조사, 부사, 감탄사 등을 제외하고 토픽모델링을 위한 명사 중심의 단어 말뭉치(corpus)를 구축하였다. 다음으로 잠재적 토픽 도출을 위하여 Blei 외(2003)가 제안한 토픽모델링 기법의 하나인 잠재 디리클레 할당, 즉 LDA 알고리즘을 적용하였으며 알고리즘의 기본 구조는 Fig. 2와 같다[20]. LDA 기법은 문서 내 토픽 분포와 키워드 생성 확률을 기반으로 하는 모델로서 문서가 내포한 잠재적 주제를 파악하는데 유용하며, 잠재적 토픽 개수(K)의 설정이 매우 중요하다[15].



**K:** number of topics,  **$\alpha$ :** proportions parameter,  **$\theta_d$ :** per-document topic proportions  **$Z_{d,n}$ :** per-word topic assignment,  **$W_{d,n}$ :** observed word,  **$\beta_k$ :** topics,  **$\eta$ :** topic parameter

Fig. 2. LDA Algorithm proposed by Blei et al(2003)

기존 연구들에서는 연구자의 주관으로 토픽 개수를 설정하였으나 최근에는 토픽 모형의 혼잡도(perplexity)와 일관성 지수(coherence score)를 검토하여 토픽의 최적 개수를 선정하고 있다[17]. 본 연구는 분석 데이터의 결과를 잘 반영하는 최적값을 선정하기 위하여 모형혼잡도와 일관성 점수를 비교하였다. LDA 알고리즘의 입력값인 토픽 수(K)를 2에서 10까지 입력하여 혼잡도와 일관성 지수를 실험하였으며, 분석 결과는 Table 1과 같다. 혼잡도가 낮고 일관성 점수는 높은 지점은 토픽 개수

Table 1. Perplexity and Coherence Score of Topic models

	Perplexity	Coherence Score
10	-6.848	0.408
9	-6.813	0.506
8	-6.83	0.395
7	-6.839	0.365
6	-6.796	0.414
5	-6.77	0.364
4	-6.714	0.373
3	-6.624	0.386
2	-6.566	0.399

6과 9지점이었으나, 본 연구의 뉴스데이터 수를 고려할 때 적합한 토픽모델의 수는 6개로 판단하여 최종적인 잠재적 토픽 수를 6으로 결정하였다.

## 4. 토픽모델링 결과

### 4.1 분석 결과

본 연구에서는 뉴스데이터 토픽모델링 분석을 통하여 6개의 잠재적 주제를 도출하였으며 각 토픽이 포함한 주요 키워드는 Table 2와 같다. 또한, 토픽별 배타적으로 포함된 주요 키워드를 중심으로 해당 토픽이 나타내는 잠재적 주제의 특성을 확인하고 토픽명을 유추하였다. LDA 기법에서 키워드는 한 토픽에만 포함되는 것이 아니라 여러 토픽에 포함될 수 있으며, 이러한 점을 고려하여 특정 토픽에만 포함된 단어를 중심으로 해당 토픽이 나타내는 주제를 예측할 수 있다[21]. 따라서 본 연구에서도 각 토픽모델에 배타적으로 포함된 키워드를 파악하고, 관련 키워드를 중심으로 한 기사를 검토하여 해당 토픽이 시사하는 잠재적 주제를 파악하였으며, Table 3과 같이 주제를 제시하였다.

Table 2. Results of LDA topic modeling

	Key words
1	<b>Bigtech.</b> Account, Network, <b>partnership.</b> User, <b>Open Banking, Mypayment. Audit.</b> Business, Pay
2	Business operator, Mydata, Utilization, Information, Custom, <b>Preliminary permit.</b> Transaction, Finance company, <b>Scope, Enforcement Decree</b>
3	<b>Amendment,</b> Bio, Signature, Data, Interest, <b>authentication,</b> Securing, utilization, Personal credit information, <b>Information and Communications Network Act</b>
4	<b>Blockchain,</b> Finance, Data exchange, Medical Information, Control, <b>Authentication, Accident, Scraping,</b> System, <b>Korea Financial Telecommunications and Clearings Institute</b>
5	Digital, <b>Pseudonym information.</b> Data, <b>Economy,</b> Activation, Utilization, Finance, Enterprise, <b>Policy,</b> Industry
6	Transaction, <b>Consortium, AI, Response,</b> Fintech, <b>Internet-only Bank, Platform, Partnership,</b> Finance, <b>Collaboration</b>

Table 3. Topic and Subject

Num	Subject of Topics
1	Digital Transformation in Finance
2	Scope of Mydata Business License
3	Amendments to Data-related Laws
4	Safe Use of Big Data
5	Data Economy Promotion Policy
6	Strategy of the Financial Industry

토픽 1은 ‘마이데이터 기반 금융혁신’의 주제를 형성하는 키워드들이 나타났다. 마이페이먼트란 새로운 금융 서비스 키워드와 함께 제휴, 페이, 네트워크, 오픈뱅킹, 비즈니스 등 빅테크 등장에 따른 금융권의 대응 관련 키워드를 함께 확인할 수 있다. 전자금융거래법 개정으로 마이페이먼트와 종합지급결제업이 가능해지며[22], 금융권의 혁신과 경쟁에 관한 논의가 확대되어왔다. ‘시중은행·빅테크, 경쟁이나 상생이나’[23], ‘금융 디지털화 가속..오픈뱅킹, 마이데이터, 마이페이먼트 대중화’[24] 등의 뉴스가 확인되었으며, 이는 전통 은행의 혁신과 빅테크의 금융업 진입에 따른 금융권 경쟁에 관한 사회적 관심을 반영한 주제라고 볼 수 있다. 토픽 2는 사업자, 마이데이터, 활용, 정보, 맞춤, 예비허가, 업무, 금융사, 범위, 시행령 등의 키워드로 마이데이터 사업 추진에 앞서 ‘마이데이터 사업자 지위 및 허가’에 관한 초기 논의로 이해할 수 있다. 현재 금융위원회에 따르면 마이데이터 본허가와 예비허가를 획득한 사업자는 총 58개 기업으로 은행, 보험, 증권 등의 금융권 기업을 비롯하여 핀테크, 빅테크 업체 등도 포함되어 있다[25]. 향후 마이데이터 사업자들의 경쟁과 차별화된 사업 영역에 관한 주제로 논의가 이동할 것으로 예상된다. 토픽 3은 개정안, 바이오, 서명, 데이터, 관심, 인증, 확보, 활성화, 개인신용정보, 정보통신망법의 키워드로 이는 ‘데이터 관련 법 개정’을 다룬 주제로 유추할 수 있다. 실제 데이터 거래와 활용을 촉진하도록 개정한 개인정보 보호법, 정보통신망법, 신용정보법, 이른바 데이터 3법에 관한 보도가 많이 이루어졌음을 확인하였다. 최근 데이터 관련 법으로는 데이터 결합과 가명정보 활용을 더욱 원활하게 하기 위한 신용정보법 시행령 및 감독규정 개정안에 관한 논의로 주제가 바뀌고 있다[26]. 토픽 4는 블록체인, 금융, 데이터 거래소, 의료정보, 통제, 인증, 사고, 스크래핑, 체계, 금융결제원과 같은 ‘안전한 데이터 활용’의 주제를 나타내는 키워드로 구성되어 있다. 금융과 의료분야는 빅데이터 활용의 주요 산업으로 주목받아왔으며, 데이터 활용의 가장 주요 이슈 중 하나는 안전한 정보 전송과 보안으로 관련 보도가 많이 이루어졌다. 블록체인, 인증, 통제, 사고, 스크래핑 등은 이와 관련한 우려와 대응방안, 데이터거래소와 금융결제원은 데이터 활용의 인증 기관과 거래 주체로서 언급되어 왔다. 2022년 1월부터는 기존의 정보 전송 방식인 스크래핑(접근 권한을 얻어 금융정보를 긁어오는 방식)이 아닌 API 방식을 통한 마이데이터 서비스가 시행된다[27]. 마이데이터 서비스가 본격화되는 상황에서 데이터의 안전한 관리와 데이터 관련 사고

가 발생할 경우에 책임소재, 기업의 대응에 관한 논의가 본격화 될 것으로 예상된다. 토픽 5는 디지털, 가명정보, 데이터, 경제, 촉진, 활성화, 금융, 기업, 정책, 산업 등의 ‘데이터 경제 활성화 정책’ 주제의 키워드로 구성되어 있다. 정부의 디지털뉴딜 정책 추진 등 빅데이터 관련 정책과 제도에 관한 뉴스가 하나의 주제로 이루어진 것으로 해석할 수 있다. 토픽 6은 업무, 컨소시엄, AI 대응, 핀테크, 인터넷전문은행, 플랫폼, 제휴, 금융, 협업 등 디지털 금융 경쟁에 대응하는 ‘사업자의 전략’에 관한 주제를 의미하는 키워드로 구성되어 있다. 금융마이데이터 서비스가 본격화되며 산업 내 플랫폼 우위를 차지하기 위한 은행, 카드, 증권, 핀테크 등 사업자 간 치열한 경쟁이 예상된 바 있다. 마이데이터 시행 이후의 최근 뉴스는 마이데이터가 디지털 금융의 격전지로서 각 사업자가 얼마나 혁신적인 서비스를 출시하는지와 사업자 간 제휴와 협업에 관한 논의가 본격적으로 이루어지고 있다[28].

#### 4.2 잠재적 주제 분석의 시사점

마이데이터와 데이터 거래를 둘러싼 잠재적 주제 분석 결과로부터 다음의 시사점을 도출하였다. 기존 논의 주제는 1) 금융권의 마이데이터 서비스 시행과 전략(토픽 1, 토픽 6), 2) 데이터 관련 법안 및 사업자허가(토픽 2, 토픽3), 3) 정부의 데이터 관련 정책과 안전한 데이터 활용(토픽 4, 토픽 5)의 사회적 이슈로 볼 수 있으며, 마이데이터의 균형 있는 발전을 위한 다음의 시사점을 도출할 수 있다.

첫째, 토픽 1과 6을 중심으로 한 금융권의 마이데이터 주제로부터의 도출한 시사점은 다음과 같다. 분석 결과에서 기존 사회적 논의가 금융 마이데이터 서비스 상용화와 금융데이터 활용에 집중된 것을 확인할 수 있다. 신용정보법 개정과 함께 금융기관이 보유한 고객데이터를 활용하는 본인신용정보관리업이 신설되고, 뉴스에서 이를 ‘마이데이터(산업)’ 혹은 ‘금융 마이데이터’로 지칭해왔다[22]. 토픽 1의 키워드들은 이러한 뉴스로부터 구성된 주제로 확인하였다. 실제로 2022년 1월 가장 먼저 마이데이터 서비스가 시행되며 뉴스 보도는 더욱 증가하였고, 토픽6의 마이데이터에서의 우위를 차지하기 위한 사업자 간 제휴와 대응 전략, 개인 고객을 확보를 위한 서비스 차별화 방안에 관한 주제로 관심이 증가하고 있다[29]. 이처럼 마이데이터의 사회적 논의가 금융분야에 집중되어 왔으나, 마이데이터 비즈니스는 의료, 통신, 생활정보, 위치 정보 등 금융권 외의 분야에서도 진행 중이다[1]. 과학기술정보통신부는 2019년부터 금융을 비롯한

의료, 에너지, 교통, 에너지, 복지, 소상공인, 생활 등 다양한 분야에서 마이데이터 실증 사업을 추진해왔으며, 2021년 마이데이터 사업 추진에 100억 원의 예산을 투입하였다[30]. 그러나 아직 실증 사업 단계로 국민 체감도는 낮고, 타 분야의 사회적 논의는 제한적인 상황이다. 향후 산업 전반의 균형 있는 발전을 위해서는 금융 외의 분야에서도 데이터 거래와 활용에 관심을 높이고 관련 논의가 이루어져야 한다. 이를 위해 정부는 금융 마이데이터 서비스를 시작으로 현재 실증단계의 사업들의 상호화를 위한 노력이 필요할 것으로 판단된다.

둘째, 데이터 관련 법안 및 사업자 허가(토픽 2, 토픽 3)의 잠재적 주제로부터 다음의 시사점을 도출하였다. 마이데이터와 같은 데이터 이동과 거래를 기반으로 한 비즈니스는 국민 개인의 데이터 주권 행사가 비즈니스 모델의 핵심이다. 개인 데이터를 토대로 비즈니스가 구축되는 경우, 사람들이 법적 권리를 행사하도록 돕는 것(또는 해당 법률을 준수하도록 돕는 것)이 매우 중요하다는 것을 이해해야 한다[31]. 그러나 기존 논의는 마이데이터의 시작과 데이터 산업 성장에 초점이 맞추어져 있고 개인의 데이터 권리에 관한 논의는 법적 해석을 중심으로만 이루어져 왔다. 정보 주체인 개인의 권리를 보장하기 위한 방향을 중심으로 주요국의 제도가 개선됨에 따라 주요 기업들은 이에 대한 대응방안을 검토하였지만[12], 지금까지는 데이터 활용을 위한 법 개정과 사업자 허가를 중심으로 한 논의에 집중되어왔다. 그러나 마이데이터 서비스가 본격화되는 시점인 만큼 데이터 활용 외에, 데이터 유출 사고 발생 시의 부정적 영향과 책임소재에 관한 법적 논의를 본격화하고 이를 사전에 방지할 수 있는 대응 전략을 수립해야 한다. 마이데이터 서비스에서는 개인의 데이터 주권 행사가 이루어지면 데이터가 기관 간 공유되고 더욱 광범위하게 이용된다. 이러한 과정에 개인, 혹은 특정 기업의 실수로 허가되지 않은 거래가 이루어지거나 정보가 유출될 수 있다. 실제로 마이데이터 서비스가 시행되자마자 정보유출 사고가 발생하였고, 기업의 미숙한 대응이 이슈가 된 바 있다[32]. 아직 법 제도적으로 마이데이터 서비스에서 개인정보가 유출된 피해자에 대한 보상이나 피해 대응 가이드라인에 관한 논의가 부족한 상황으로, 사업자가 더욱 적극적으로 개인데이터 유출의 위험에 대응하도록 하는 방안이 필요하다. 또한, 지금까지 일반 대중은 데이터에 관한 프라이버시나 유출위험에 관하여 큰 인식이나 반대 없이 정보 제공 동의를 해왔지만, 마이데이터 서비스에서는 데이터의 가치에 관한 고객의 인식이 변화할 것이다. 따라서 데

이터 사고가 발생하는 경우 마이데이터 사업자의 비즈니스 모델에 부정적인 영향을 미칠 수 있다. 따라서 정보 주체인 개인 고객의 정당한 권리와 이익을 보장하고, 안전한 마이데이터 서비스가 제공될 수 있도록 데이터 관련 사고와 관련한 대응방안에 관한 구체적인 논의가 필요하다.

셋째, 정부의 데이터 정책과 안전한 데이터 활용에 관한 주제(토픽 4, 5)로부터 다음의 시사점을 도출하였다. 지금까지 정부의 마이데이터 관련 정책(토픽5)은 산업활성화 정책과 데이터의 유통을 촉진하기 위한 내용을 중심으로 추진됐다. 안전한 데이터의 활용(토픽 4) 또한 마이데이터의 보안을 위한 기술적인 방식은 무엇으로 할지 등의 논의에 집중되어 있다. 최종적으로 금융마이데이터는 API를 의무화하며 보안을 강화했으나, 실제 서비스 개시 이후 오히려 보안에 관한 우려가 더욱 커지고 있다[33]. 한국데이터산업진흥원의 2019년 마이데이터 현황조사 결과에 따르면, 개인정보 활용에 대하여 기관에 대한 신뢰가 있는 경우 제공 의사가 높아졌고 이러한 신뢰를 판단하는 요소로 보안이 나타났다[5]. 결국 데이터 거래 시장 활성화를 위해서는 기업 간 데이터 공유가 안전하게 이루어져야 하는 것을 전제로 하고 있으며, 고객데이터의 안전을 보장하고, 사업자에 대한 신뢰를 확보하는 것이 비즈니스 활성화의 핵심적인 요인이다. 종합하면, 데이터 활용 촉진을 위해서 정부는 기업 간 안전한 데이터 거래 환경을 위한 구체화 된 보안정책과 거버넌스를 마련해야 하며, 정보의 비식별화와 인증제도 외에도, 데이터를 보호하기 위한 기술과 인프라 구축, 향후 서비스 사용자가 증가하며 발생 가능한 관련 사고에 대응하는 구체화 된 정책에 관한 논의가 필요하다.

Table 4. Potential Topics and Implications

Categorization	Implications
My data service implementation and strategy in the financial sector	<ul style="list-style-type: none"> <li>Limited discussion on My Data in other fields</li> <li>The necessity of discussing data transactions and utilization in various fields for the balanced development of the industry</li> </ul>
Data-related laws and business licenses	<ul style="list-style-type: none"> <li>Discussion focused on amendments to laws for data use and business licenses</li> <li>The need for a full-fledged legal discussion on the negative impact and liability in the event of a data breach</li> </ul>
Government data-related policies and social issues of safe data use	<ul style="list-style-type: none"> <li>Discussion focused on industrial revitalization policies and ways to promote data distribution</li> <li>The need to discuss specific security policies and governance for establishing a safe data transaction environment between companies</li> </ul>

## 5. 결론

본 연구는 마이데이터가 본격화되는 현시점에 데이터 활용과 마이데이터에 관한 사회적 논의 양상이 어떻게 이루어지는지 파악하여 잠재적 주제를 도출하고자 하였다. 뉴스데이터를 LDA 토픽모델링 기법을 통하여 뉴스 데이터를 분석하였으며, 향후 빅데이터 활용과 거래 활성화를 위한 주요 시사점을 다음과 같이 제시하였다.

기존 마이데이터를 둘러싼 사회적 논의의 주제는 금융권의 마이데이터 서비스, 데이터 활용에 관한 초기 논의로서 데이터 활용을 촉진하고 산업을 성장시키기 위한 정책과 제도, 데이터의 상업적 목적의 이용을 위한 법·개정안, 마이데이터 서비스 사업자 허가, 보안 기술 등의 사회적 이슈를 중심으로 한 잠재적 주제가 형성되어 있음을 확인하였다. 본 연구에서는 금융에 집중된 마이데이터에 관한 사회적 논의가 의료, 교통, 소비 등 다양한 산업으로 확장되고, 개인 정보 침해, 사고 대응을 위한 법제도와 보안정책에 관한 논의가 더욱 구체화 되어야 한다는 시사점을 도출하였다. 기존 연구에서도 마이데이터 사업이 활성화를 위한 중요한 요인으로 보안정책과 프라이버시 유출 문제 해결 등을 제시한 바 있다[6].

마이데이터가 본격화되며 국내 데이터 경제기반 마련에 정부와 산업의 관심이 증가하고 있으며, 실증 사업으로 추진되어 오던 금융마이데이터 서비스가 시작되면서 사업의 성공 여부에 관심이 집중되고 있다. 서비스 개시 이후 정보유출 사고가 발생하는 등 논란으로 서비스의 성공에 관한 우려가 커진 상황이지만[33], 데이터거래의 활성화와 데이터를 통한 산업혁신을 보여주는 데이터 경제의 중요한 비즈니스 사례가 될 것이라 기대는 계속되고 있다. 그러나 기존 마이데이터 연구들은 법제도적 논의에 집중되어 있었으며[10,11], 초기 연구로서 마이데이터 생태계[1,13]와 사업 모델을 제시하였으나 종합적인 관점에서 살펴본 연구는 아직 부족하였다. 따라서 본 연구는 데이터 3법 개정 및 마이데이터 시행과 함께 데이터 거래와 데이터 활용에 관한 논의의 주제가 어떻게 이루어져 왔는지 종합적인 시각에서 살펴보았다는 의의가 있다. 또한, 기존 논의로부터 잠재적 주제를 분석하여 향후 논의가 이루어져야 할 공백을 파악하여 시사점을 제시하였다. 그러나 본 연구 결과는 서비스 개시 이전의 뉴스데이터를 분석하였다는 점에서 실제 서비스 제공과 데이터 거래 과정에 나타날 수 있는 사회적 이슈와 주제를 반영하지 못하였으며, 뉴스데이터를 분석하였기에 실제 이용자와 사업자의 인식을 구체적으로 파악하지 못한

한계가 있다. 따라서, 후속 연구에서는 데이터 수집 기간을 마이데이터 서비스 개시 이후로 확대하여 기간별 결과를 비교한다면 더욱 유용한 시사점을 제시할 수 있을 것이다. 또한, 실제 마이데이터 서비스가 시행된 만큼 이용자를 대상으로 한 실증 연구가 진행될 필요가 있다. 기존 연구는 서비스 시행 이전에 법리적 논의에 집중되어 있어 이용자 관점의 연구는 제한적이었다. 따라서 향후 연구는 이용자 증가와 함께 나타날 사회적 이슈와 사용자 인식에 관한 연구가 이루어지길 기대한다. 또한, 실제 마이데이터 서비스 운영과정에 사업자가 직면하는 어려움에 관한 구체적인 분석과 함께 금융, 의료, 제조 등 각 산업별 마이데이터가 가지는 특징 등을 세밀하게 분석한다면 더욱 유용한 결과를 제시할 수 있을 것으로 기대한다.

## REFERENCES

- [1] K.R Yang, S. K. Park & B.G. Lee. (2021). The MyData Business Ecosystem Model. *Journal of Digital Convergence*, 19(11), 167-180. DOI:10.14400/JDC.2021.19.11.167
- [2] J.S. Jeon & H.Y. Kwon. (2018), Comparative Analysis of 4 Countries' Legal System Regarding Personal Information, Pseudonymised Information, and Anonymous Information, *Journal of Korea Information Law*, 22(3), 183-218.
- [3] D.H. Bang. (2018). A Study on Enactment of Data Legislation for Activating the Data Economy. *Law Review*, 59(1), 77-104. DOI:10.35275/pnulaw.2018.59.1.003
- [4] D.H. Min & J.S. Oh. (2018). *A plan to activate data transactions for the development of ICT-based new industries*. Korea Information Society Development Institute. Seoul
- [5] Korea Data Agency(2019), *2019 My Data Status Survey Report*, Seoul.
- [6] J.K. Bae. (2021) A study on the Legal and institutional factors for activation of MyData Industry, *Logos Management Review*, 19(1), 2021.3, 117-132. DOI:10.22724/LMR.2021.19.1.117
- [7] S.K. Park, H.J. Lee. & B.G. Lee. (2021). Exploring Social Issues of On-demand Delivery Platform Participants. *Journal of Digital Convergence*, 19(7), 79-85. DOI:10.14400/JDC.2021.19.7.079
- [8] S.K. Kim, S.J. Lee. & J.G. Kim (2016). A Study on the Development of Phased Big Data Distribution Model Based on Big Data Distribution Ecology. *Journal of Digital Convergence*, 14(5), 95-106. DOI:10.14400/JDC.2016.14.5.95

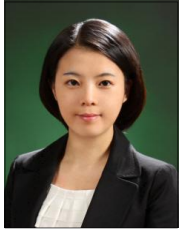
- [9] Cao, X., Chen, Y., & Liu, K. R. (2017). Data trading with multiple owners, collectors, and users: An iterative auction mechanism. *IEEE Transactions on Signal and Information Processing over Networks*, 3(2), 268–281.  
DOI:10.1109/TSIPN.2017.2668144
- [10] J.H. Park. (2020). Activation of My Data System and Legal Issues. *Ajou Law Research*, 14(1), 96–119.
- [11] J.M. Choi. & Y.E. JO. (2020), A Study on the Right to Data Portability and "MyData" Industry, *Journal of Law & Economic Regulation*, 13(2), 92–107.  
DOI:10.22732/CELP.2020.13.2.92
- [12] M.J. Song & I.S. Kim (2019). A Study on Privacy Protection in Financial Mydata Policy through Comparison of the EU's PSD2. *Journal of the Korean Institute of Information Security and Cryptology*, 29(5), 1205–1219.  
DOI:10.13089/jkiisc.2019.29.5.1205
- [13] H.J. Park. & J.H. Yang (2020). A study on the establishment of Health MyData ecosystem in the public domain, *The Journal of Korea Institute of Information, Electronics, and Communication Technology*, 13(6), 511–522  
DOI:10.17661/jkiect.2020.13.6.511
- [14] H.T. Oh. & J.H. Yang (2020). Personal Health Record/Electronic Medical Record Data Trading Model for Medical My Data Environments, *The Journal of Korea Institute of Information, Electronics, and Communication Technology*, 13(3), 250–261.  
DOI:10.17661/jkiect.2020.13.3.250
- [15] J.Y.An, K.B. Ahn. & M. Song (2016), Text Mining Driven Content Analysis of Ebola on News Media and Scientific Publications, *Journal of the Korean Society for Library and Information Science*, 50(2), 289–307.  
DOI:10.4275/kslis.2016.50.2.289
- [16] van Altna, A. J., Moerland, P. D., Zwinderman, A. H., & Olabbarriaga, S. D. (2016). Understanding big data themes from scientific biomedical literature through topic modeling. *Journal of Big Data*, 3(1).  
DOI:10.1186/s40537-016-0057-0
- [17] S.K. Park. & B.G. Lee. (2019), A Text Mining Approach to the Analysis of Issues for Accommodation Sharing Business, *Journal of Tourism and Leisure Research*, 31(7), 209–229.  
DOI:10.31336/jtlr.2019.7.31.7.209
- [18] S.H. Noh. (2020). Analysis of Issues Related to Artificial Intelligence Based on Topic Modeling, *Journal of Digital Convergence*, 18(5), 75–87.  
DOI:10.14400/JDC.2020.18.5.075
- [19] S.S. Lee, I.H. Yoo. & J.H. Kim. (2020). An analysis of public perception on Artificial Intelligence(AI) education using Big Data : Based on News articles and Twitter. *Journal of Digital Convergence*, 18(6), 9–16.  
DOI:10.14400/JDC.2020.18.6.009
- [20] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). "Latent dirichlet allocation", *Journal of machine Learning research*, 3, 993–1022.  
<https://www.jmlr.org/papers/v3/blei03a>
- [21] S. K. Park. & B. G. Lee. (2019). A Topic Modeling Approach to the Analysis of Issues for Kids Creators. *Culture and Convergence*, 41, 549–578.  
DOI:10.33645/cnc.2019.08.41.4.549
- [22] H.J. Noh.(2021), *MyData Business Status and Insurance Company Implications*, Korea Insurance Research Institute. Seoul.
- [23] S.I. Back. (2021.05.14.). *Commercial Banks:Big Tech: Competition or Win-Win relationship?*. Global Economic.  
[https://news.g-enews.com/ko-kr/news/article/news\\_all/202105141534567766da878e5c53\\_1/article.html?md=20210514154604\\_S](https://news.g-enews.com/ko-kr/news/article/news_all/202105141534567766da878e5c53_1/article.html?md=20210514154604_S)
- [24] H.K. Kim. (2021.10.15.). *Accelerating financial digitization...Open banking, MyData, MyPayment became popular. Apple Economy.*  
<http://www.apple-economy.com/news/articleView.html?idxno=68122>
- [25] M.H. Noh. (2021.12.14.). *[mydata reset]③ Look, it's a big data battle.Business watch.*  
<http://news.bizwatch.co.kr/article/finance/2021/12/13/0010>
- [26] C. L. Ha. (2022.01.06.). *Even startups that do not have big data can easily combine and utilize pseudonymous information.* Yonhap News.  
<https://www.yna.co.kr/view/AKR20220106083900002?input=1195m>
- [27] K.N. Lee. (2021.12.10.). *[mydata reset]① What does "My information is mine" mean?*. Business watch.  
<http://news.bizwatch.co.kr/article/finance/2021/12/09/0025>
- [28] J.H. Shin. (2022.01.06.). *"Catch MyData, the financial secretary in your hand."*, *The life or death battles of financial companies.* Dong-A ilbo.  
<https://www.donga.com/news/article/all/20220105/11113419/1>
- [29] J.W. Suh. (2022.01.04.). *A financial company that collects customer information, a fintech that learns know-how...My data do make strange bedfellows.* Maeil Business Newspaper.  
<https://www.mk.co.kr/news/economy/view/2022/01/10139/>
- [30] M.C. Lim. (2021.03.18). *Discovering My Data Services in the Financial, Medical, Consumption, and Transportation Fields.* Aju Business Daily.  
<https://www.ajunews.com/view/20210318195405416>
- [31] Maldonade, S.(2019), *Mydata Business Models.* December 31, 2019,  
<https://privacycloud.com/en/mydata-business-models/>
- [32] J.H. Kim.(2022.01.05.). *MyData information leakage.* Etnews. <https://www.etnews.com/20220105000161>



- [33] Herald Economy (2022.01.05), *My Data D-Day... "security" anxiety of Small fintech*. Herald Economy. <http://news.heraldcorp.com/view.php?ud=20220105000221>

조 지 연(Ji Yeon Cho)

[정회원]



- 2018년 : 연세대학교 정보시스템 박사
- 2018년 9월 ~ 현재 : 연세대학교 방송통신정책연구소 전문연구원
- 2019년 3월 ~ 현재 : 연세대학교 객원 교수
- 관심분야 : 빅데이터, 데이터 비즈니스, IT정책

· E-Mail : jy.cho@yonsei.ac.kr

이 봉 규(Bong Gyou Lee)

[정회원]



- 1988년 : 연세대학교 상경대학(학사)
- 1992년 : Cornell University(석사)
- 1994년 : Cornell University(박사)
- 2005년 ~ 현재 : 연세대학교 정보대학원 교수
- 2018년 ~ 2020년 : 연세대학교 학술정보원 원장

· 관심분야 : 플랫폼 비즈니스 전략, ICT·콘텐츠, IT정책  
· E-Mail : bglee@yonsei.ac.kr