

Using big data on women and family for policy development and improvement(Ⅲ)

Sung Mi Jung, You Kyoung Moon, Soo Yeon Lee,
Ji So Yoon, Yeon Gyu Lim, Song Yi Park, Jae Kyung Lee



**Using big data on women and family for
policy development and improvement (III)**

©2020

Korean Women's Development Institute

225 Jinheung-ro Eunpyeong-gu

Seoul, 03367, Republic of Korea

www.kwdi.re.kr

Contents

I . Introduction	1
II . Policy and Research Trends of Big Data on Women and Families ..	3
III . Comparison and Analysis of Gendered Political Debate in Online Communities	7
IV . Empirical Analysis and Proposed Countermeasures on the Reproduction of Gender Discrimination in Big Data Algorithms	9
1. Experimental Empirical Analysis of Potential Gender Discrimination in Big Data Algorithms	9
2. Institutional Suggestions for Resolving Gender Discrimination in Big Data Algorithms	14
V . Infrastructure and Governance Strategies for Promoting the Utilization of Big Data for Women and the Family	16
1. Survey on the Present Conditions and Perceptions of Big Data on Women and Families	16
2. Proposals for the Establishment of Big Data Platforms for Women and the Family	17
3. Suggested Directionality for Building Governance of Big Data on Women and the Family	18

VI. Conclusion and Policy Proposals	21
1. Proposed Research for Big Data on Women and the Family ...	21
2. Revision of Regulations for Processing Personal Information ...	23
3. Constructing Governance for Big Data on Women and the Family	25
Appendix 1.	28

Tables

〈Table 1〉	Examples of domestic big data platforms	5
〈Table 2〉	Examples of foreign big data platforms	6
〈Table 3〉	Distribution of gendered recruitment rates	10
〈Table 4〉	Gendered Employment Rates of Algorithms based on Gender-biased Recruitment Data	10
〈Table 5〉	Gendered Employment Rate of Algorithms Learned from Data w/out Gender-biased Data	12

Using big data on women and family for policy development and improvement (III)¹⁾

Sung Mi Jung

You Kyoung Moon

Soo Yeon Lee

Ji So Yoon

Yeon Gyu Lim

Song Yi Park

Jae Kyung Lee

I . Introduction

- As computerized industries develop and diversify across all fields, the industrial and political value of big data in the Fourth Industrial Revolution continues to accumulate. Domestic and foreign public sectors are deriving new insight and value from big data in the process of data-driven decision-making. As such, big data is currently being actively utilized in the fields of public health and welfare, culture and tourism, transportation, national defense, and education. With the concurrent development of complementary technologies such as cloud computing, IoT, and machine learning, the utilization of big data is expected to become even more cost-effective, accessible and innovative in the future.

1) This research is a comprehensive report of the National Research Council for Economics, Humanities and Social Studies Cooperative Research Studies, and includes the contents of two designated research efforts. See Appendix 1 for more information.

-
- We expect that the utilization of big data to establish and implement policies regarding women and families will allow for the introduction of new research topics and results previously unaddressed by traditional methods. While the existing research conducted on the topic is meaningful in that it focused on the possibilities for big data utilization, there is as of yet a marked lack of research on the actual methods or policies of usage that would lead to promotion of big data usage.
 - In this study, we intend to explore the future possibilities for using big data in women and family policies by discussing the mid-to-long-term potential of establishing a platform for the utilization of big data for women and families as well as an appropriate form of governance. To be more specific, we review the possibility of using applicable data to construct a usable platform, and also examine the establishment of governance measures necessary for the utilization of big data. For a pilot analysis, we first used social network data to analyze discourse on gender politics in online communities, and secondly reviewed the possibility of gender discrimination and perpetuation of existing gender discrimination by big data algorithms.
 - To that end, we reviewed a wide breadth of domestic and international literature and data to amass examples of big data used for the benefit of women and families in the private and public sectors. Furthermore, hybridized research efforts spanning between various industries and academia were commissioned to relevant institutions to promote research collaboration between experts in various fields. The proposals for these delegated research efforts were composed after approval by the internal Research Review

Committee/Board. The joint research was conducted in stages, namely a start-up briefing, an interim briefing, and a result briefing intended to consolidate efforts and seek solutions between the various research agents and institutions. The topics of analysis included first empirical analysis and experimental research on improvement measures for the reproduction of gender discrimination in big data algorithms, and the infrastructure and governance construction strategies for promoting the utilization of big data for women and families. Each study was conducted by the Korea University Labor Affairs Research Institute and the Institute for Information and Communication Policy.

II. Policy and Research Trends of Big Data on Women and Families

- The phenomenon of new data production through data linkage has emerged as a prominent issue in big data research, generating new implications and leading to extensive research efforts across various fields. Therefore, inter-data connectivity and linkage has become an important point of practical consideration in research. Related research efforts have mainly focused on identifying methodological issues in data linkage and combination, supplementing the limits of combined data to produce more complete results, and examining legal and institutional issues in data linkage such as the potential for violations of personal information and privacy.
- The practical problems and solutions related to the methods and results of data linkage are relatively likely to be addressed through

the improvement of collation processes and statistical methodology. However, we can still observe several pertinent legal and institutional issues in data linkage. Most prominently, various inter-data linkages increase the risk of personal identification while also presenting potential fixes to such issues. As a result, related government bureaus such as the Office for Government Policy Coordination and the Ministry of Government Administration and Home Affairs have published 'Guidelines for Non-identification of Personal Information' in June 2016 as an attempt to protect personal information in the face of increasing big data usage but as of yet this measure does not present any viable legal or normative effects.

- A revision to the current Personal Information Protection Act has been proposed to supplement its policies with regards to big data, and efforts have been made to amend systems of identity protection and implement non-identification measures through statistical methods. In this vein, issues in governing big data are among the most vital points of contention for the construction and utilization of big data platforms in women and family policies.
- Recently, the Public Data Strategy Committee established the “2019 Implementation Plan for the Provision and Utilization of Public Data” (2019.2.28) as part of the the third year of its “2nd Strategy for the Provision and Utilization of Public Data” ('17~'19). The committee announced several major projects it was undertaking to promote the widespread usage of big data, such as expanding the availability of public data, improving access through systematic management of data, and promoting the active utilization and analysis of public data.

○ Examples of domestic and foreign big data platforms are organized in the following tables

〈Table 1〉 Examples of domestic big data platforms

Department/Ministry	Platform	Provided information/services
Ministry of Public Administration and Security	Open Data Portal (www.data.go.kr)	<ul style="list-style-type: none"> - Analytical data/services such as file data and open API for public data of public institutions - Analysis of data visualization - Usage cases - Forums for analysts to share information
Korea Communications Commission	Media Statistics Portal (www.mediastat.or.kr)	<ul style="list-style-type: none"> - Statistical results for the eight fields examined by the Korea Communications Commission
Ministry of Science and IST (MSIT)	IT Statistics of Korea (ITSTAT) (www.itstat.go.kr)	<ul style="list-style-type: none"> - Time series values of ICT statistics classified by subject - Time series data: statistical graphs, tables, commentary, publications - Key macro indicators - Data such as statistical notifications and reports
Ministry of Culture and Sports	Culture Portal (www.culture.go.kr)	<ul style="list-style-type: none"> - Information on Culture and Arts, Culture Industry, Tourism, Cultural Heritage, Sports, Literature in Korea - Information on cultural experiences (performances, exhibitions, events, etc.) and ticket sales - Cultural communications (cultural issues, cultural videos, etc.) - Content related to cultural knowledge (culture spaces, Top 100 Korean culture artifacts, etc.) - News (policy, recruitment information, etc.) - Enables industrial usage (traditional motif DB, 3D contents, open API)
National Statistical Office (NSO)	Korean Statistical Information Service (KOSIS) (www.kosis.kr)	<ul style="list-style-type: none"> - Statistical charts of domestic approval statistics organized by subject and organization - International statistics, North Korean statistics - Visualized analysis contents on relevant issues - Online publications such as statistical reports
National Research Council for Economics, Humanities & Social Science	Future Prediction and Policy Support System (Planned for construction)	

Source: Each platform site was accessed and summarized by authors (Access date: 2019. 9. 27.)

〈Table 2〉 Examples of foreign big data platforms

Institution	Platform	Provided information/services
UN	Big Data UN Global Working Group (https://unstats.un.org/bigdata/)	<ul style="list-style-type: none"> - Website currently shares 184 cases in the inventory of big data projects - Introduces various research topics, including economy, employment, transportation, agriculture, and health care, focusing on the eight tasks of the UN GWG.
EU	European Data Portal (https://www.europeandataportal.eu/en)	<ul style="list-style-type: none"> - Currently 926,583 data points available - Data in 13 areas: agriculture and fishing industries, energy, rural and urban areas, transportation, economy and finance, international issues, public sector, law, environment, education, health, population, science and technology
USA	DATA.GOV	<ul style="list-style-type: none"> - Currently 236,236 data points available - Data in 14 areas: agriculture, climate, consumerism, ecosystem, education, energy, finance, health care, local government, manufacturing, shipping, marine, public safety, science and research
UK	DATA.GOV.UK	<ul style="list-style-type: none"> - Currently 52,771 data points available - Data in 12 areas: business and economy, crime and justice, national defense, education, environment, government, government expenditure, health care, map mapping, society, urban and rural areas, transportation

Period: Data accurate as of September 2019

Source: Each platform site was accessed and summarized by authors (Access date: 2019. 9. 27.)

III. Comparison and Analysis of Gendered Political Debate in Online Communities

- This study aims to understand the main political discourses of respective female and male dominant online communities and how the two compare with each other. After selecting representative male-centric communities A, B, C and female-centric communities A, B, and C for analysis, posts on specific bulletin boards of individual sites were collected through web crawling techniques over a period of over six months, and then analyzed for text.
- This study reveals the distinct differences in gendered political discourse through topical analysis and semantic network analysis (SNA) of political postings in online communities. First, a comparison between the male community A and the female community A has demonstrated that gender-based ‘hate speech’ prevails within political discourse in both communities. Further comparisons of other male and female communities then make clear the main differences in how the different genders conduct online political discourse. Other male (community B, C) / female (community B, C) communities show a similar tendency to discuss ‘formal’ politics such as the economy, party politics and crime, yet differ in their approaches. For example, when discussing a single ‘criminal incident’, the discourse within these male communities tended to focus on the politics of the investigation, while the corresponding female communities would focus on the sex trafficking involved in the investigation. In addition, examining the topics that were unique to male and female communities also produced interesting results, with women mainly discussing ‘social’

political topics such as childcare, education, safety and natural disasters, while men focused on specific ‘formal’ political topics such as North Korea, elections and democratization. As such, we could observe that while men are more interested in formal political issues than women, women are more interested in the issues that relate closely to daily life, including highly gendered issues such as childrearing and caregiving. Both genders also demonstrate an activist tendency to formalize their key issues by starting public petitions.

- In conclusion, this study proves that online communities are significant spaces that foster everyday political discourse. It further shows that the nature of the discourse within those spaces varies depending on the gender characteristics of the community and that in the real of politics, policy needs may vary significantly between genders. We expect that the methodology used in this study can be extrapolated and applied to other online communities in the future to identify trends in gendered opinions on major social issues such as low birth rates, Korean unification and security, and economic policies (e.g., minimum wages, housing policies). However, when taking potential research ethics issues into account, future research efforts will need to make use of communities that are relatively easy to access and open to the general public at large, as well as prepare a platform for collecting and archiving posts indefinitely rather than temporarily for the sake of analyzing and tracking the trends of discourse as they happen over extended periods of time.

IV. Empirical Analysis and Proposed Countermeasures on the Reproduction of Gender Discrimination in Big Data Algorithms

1. Experimental Empirical Analysis of Potential Gender Discrimination in Big Data Algorithms

A. Purpose and methodology of research

- This study was intended to empirically confirm existing concerns that big data algorithms reproduce and amplify patterns of prejudice and discrimination, as well as to find methods of eliminating or mitigating such reproductions of discrimination.
- In order to achieve the above research objectives, the study specifically examined the subject of employee recruitment algorithms, referring to employment audit research methods to collect data through an experimental recruitment evaluation survey, developing a novel recruitment algorithm through guided machine learning methods, and analyzing and evaluating the predictive results of the algorithm.
- Based on the existing patterns of hiring and employment for programmers with significant work experience in the SI industry, researchers generated resumes for virtual applicants with equivalent qualifications across four demographic groups differing in gender and childed status, and collected data by surveying hiring personnel currently active in the industry.

B. Confirmation of gender discrimination in recruitment practices

- Our study confirmed the existence of gender discrimination in the

processes of recruiting and hiring experienced programmers in the IT industry's information system integration service (SI industry).

- We observed a pattern of systematic sexism in recruitment practices that preferentially hired more men than women, despite all other conditions being equal.

〈Table 3〉 Distribution of gendered recruitment rates

	Male	Female	Total
Rate of recruitment	23%	16%	20%

C. Confirmation of reproduction of gender discrimination in recruitment algorithms

- These results confirm that the recruitment algorithm, which was developed from learning data with inherent gender biases, perpetuated sexist recruitment and hiring practices by further reproducing and strengthening existing patterns of discrimination.

- We can infer that machine learning methods inadvertently strengthen such patterns of sexism and discrimination due to the inherent tendency of machine learning to recognize and replicate existing trends in the given data.

〈Table 4〉 Gendered Employment Rates of Algorithms based on Gender-biased Recruitment Data

Method of development	Male	Female	Total
Logistic regression	17%	10%	13%
Support vector machine	14%	9%	11%
Artificial neural network	18%	9%	14%
Boosting	19%	11%	15%
Random Forest	16%	11%	14%

D. Confirming the elimination of gender discrimination in recruitment algorithms

- Results confirm that the use of employment algorithms developed without the basis of potentially discriminative data such as gender and childed status fundamentally circumvents gender biases in recruitment and ultimately promotes gender-equal employment.
 - From these results we can infer that there was no significant difference between the genders when it came to the non-gendered key variables that affect employability. In other word, there was no other factor other than gender or highly gendered variables that correlated to the disparity in hiring. We further believe these results were possible because the data in this study was laboratory-generated and the conditions of all inputs were completely equivalent by gender. In other words, eliminating gender-related learning data can create a recruitment algorithm that is much more equitable on the basis of gender.
 - There was little difference in overall predictive accuracy between algorithms learned from all data and algorithms that excluded data with discriminatory properties.
 - Exclusion of gendered data resulted in better predictions for female recruitment, but somewhat lower predictive accuracy for male recruitment.
 - It is not always correct or necessary to develop and evaluate algorithms specifically to mitigate discrimination of a marginalized group, but such measures can be appropriate depending on the circumstances.

〈Table 5〉 Gendered Employment Rate of Algorithms Learned from Data w/out Gender-biased Data

Method of development	Male	Female	Total
Logistic regression	13%	13%	13%
Support vector machine	12%	12%	12%
Artificial neural network	13%	13%	13%
Boosting	13%	15%	14%
Random Forest	14%	14%	14%

E. Identifying disparities between methods of developing recruitment algorithms

- In terms of developing recruitment algorithms, there was some disparity in the predictive outcomes for employment.
 - This difference most likely depends on the subject of analysis, the characteristics and amount of data, and the characteristics of the input variables.
 - Therefore, it is effectively not possible to generalize the value of development methods for specific topics or assume the differences between development methods are negligible, but it can be appropriate to develop various algorithms and subsequently select and utilize the most appropriate development model through comprehensive algorithm evaluation.
- White box development methods such as logistic regression from the recruitment algorithm development method provide information on the factors that affect the predicted results without sacrificing efficacy of the algorithm performance.
 - Therefore, it is necessary to develop algorithms through various methods and finalize appropriate methods by collectively considering

the predictive accuracy, demographic equity, and comprehensibility (transparency) of the results.

F. Significance and limitations of research

○ The significance of this study is as follows:

- Gender discrimination in employment and hiring practices can essentially be confirmed through employment audit methods.
- Identification of detrimental concerns such as the possibility of sex discrimination in big data algorithms through empirical research.
- Empirically confirmed that algorithms, specifically employability screening algorithms, should not contain any personal or protected information that can be perceived as discriminatory.
- Identifies the need for research on algorithm auditing in the AI era to avoid reproducing social discrimination such as gender biases and provides rationale for expanding such research.

○ The limits of this study are as follows:

- There are inherent limits to the value of laboratory-generated data. This study calls for further development into research based on actual employment data.
- Gender discrimination in employment is essentially confirmed through employment audit methods. However, simply excluding potentially gender-biased personal data is not sufficient to effectively deter discrimination. It is necessary to conduct further research on the possible reproductions of gender discrimination in algorithms with regards to various forms of proxy personal information.

- Furthermore, it is necessary to conduct an algorithm audit study on algorithms that are being actively used to identify whether they reproduce gender discrimination. Such a project should first prioritize public institutions such as the government.

2. Institutional Suggestions for Resolving Gender Discrimination in Big Data Algorithms

- Three core principles of fairness/fairness, accountability and transparency are necessary to establish the ethics of big data algorithms and counter the negative impact of big data algorithms, while an "audit system" has been proposed as one of the more practical means of overcoming such incipient detrimental effects.
- The algorithm is defined in the following terms:
 - An algorithm cannot be understood as a simple code or formula but rather is comprised of several inter-connected subsystems. As a system that utilizes machines as observers that make multiple decisions in conjunction with human intervention, it can be defined as technical governance factors combined with aspects of control that operate to achieve a particular goal, in addition to simple logical circuits or a set of elemental rules. Therefore, it is necessary to attempt to offset the potential negative effects of such a system by identifying the social factors and circumstances surrounding the objective data and human influence that compose algorithms.
- Relationship of the three principles of fairness, accountability and transparency (FAT)
 - The relationship between the three principles of fairness,

accountability, and transparency can be defined as such, First, in order to adequately counter the human social (negative) effects of algorithms, overall algorithm accountability is necessary. Second, refraining from perpetuating injustice, especially through the reproduction of discrimination and prejudice, is among the most vital and fundamental purposes of accountability within algorithms. Third, transparency is constitutionally important as both a means and condition of ensuring accountability.

- In Korea, there is no actual audit research on algorithms being conducted at the FAT level and such action is long overdue.
- In order to prevent gender discrimination in algorithms, it is necessary to monitor all phases of planning, design, development and operation of algorithms by the standards of FAT.
- However, any immediately possible steps towards progress should be taken. The elimination of gender-indicating information in learning data reduces gender discrimination in the function of the algorithm. This should be appreciated on its own merits.
- It will be necessary for a third-party audit agency to conduct an audit on algorithms used by national and local governments (Seoul, Gyeonggi-do, etc.), particularly with relation to gender and female-related topics. The Korean Women's Development Institute expects to be able to contribute to that end.

V. Infrastructure and Governance Strategies for Promoting the Utilization of Big Data for Women and the Family

1. Survey on the Present Conditions and Perceptions of Big Data on Women and Families

- The Policy Research Information Service & Management (PRISM) was used to review the status of survey statistics for women and the family, revealing that a total of 689 studies were ordered by the Ministry of Gender Equality and Family between 2001 and September 2019, 659 of which have been publicly reported and managed. Out of these, the Ministry of Gender Equality and Family's research services, excluding child and youth research, totaled 215 cases between 2013 and 2018, with 122 in the field of gender equality, 39 in human rights protection and 48 in the family sector.
- The 2015-2018 survey statistics on women and the family (nationally approved statistics) totaled 11 cases, excluding youth-centric statistics and surveys.
- In addition, members of the Ministry of Gender Equality and Family, related agencies, and other data experts were interviewed and surveyed regarding the promotion of management and utilization of big data, and the infrastructure and improvements of big data analysis, revealing that there is no current plan for big data production and management within the Ministry of Gender Equality and Family. There are no personnel in charge of big data-related tasks in the ministry's organization, and statistics on women and the

family are overseen and managed by the Information Statistics Department under the Planning and Coordination Office.

- Currently, most of the data being managed within the Ministry of Gender Equality and Family is at the level of open public data and statistical support, and it is difficult to say that such data is being managed systematically or in comprehensively.
- Idea for potential improvement include promoting awareness of the significance of data, increase of manpower, using data to find relevant policy subjects, and improving data accessibility.

2. Proposals for the Establishment of Big Data Platforms for Women and the Family

- Big data efforts in the women and family sectors lack sufficient data to build upon. Under these circumstances, the following plan to big data platforms for women and the family is derived from domestic and international cases of big data platform construction: 1) Establishment of a new big data platform for women and the family; 2) Linkage and amendment of existing big data platforms for women and the family
- A big data platform could be established to enable higher-level analysis by linking general individual-level data related to women and the family to data from more specific regions, institutions, etc.
- The establishment of a platform for big data on various topics relating to women and families would contribute to the further discovery and propagation of relevant and practically useful public and industrial innovation services beyond big data collection,

analysis, and distribution, including linking distributed data, and the eventual creation of a data ecosystem for women and families based on big data.

3. Suggested Directionality for Building Governance of Big Data on Women and the Family

A. Parameters and significance of data governance

- Data Governance generally refers to policies and processes used to manage the availability, usefulness, integration, and security of data used by enterprises and emphasizes compliance with privacy, security, data quality, and management regulations (Jung Yong-chan, 2018. Original source: Laudon and Laudon, 2011). In recent years, this definition further includes the exercise of authority, control and shared decision-making in data asset management (Jung Yong-chan, 2018. Original sources: Ladley, 2012).
- The environment for statistical survey production is deteriorating despite the growing importance of data. As such, data policies in major countries around the world are suggesting that data strategies should be a key subject in government innovations.
- The importance of data governance in a big data environment is highlighted by the fact that the design and allocation of authority within the processes of obtaining, sharing and utilizing data is vital, particularly when converging and analyzing data from various sources (Lee Ki-joon et al, 2019).

B. Cases of reformed data governance

- The UK published their “Government Transformation Strategy 2017-2020” report wherein they highlight the appointment of a Chief Data Officer in the government to lead the use of data and enacted comprehensive information access for government departments through the revision of the Digital Economy Act 2017 in order to appropriately prepare national statistics.
- The U.S. has published the report "The Promise of Evidence-Based Policymaking, 2017", which calls for setting up a 2016 Evidence-Based Policy Committee to strengthen administration based on data and innovatively improve the way government uses and protects collected data. Furthermore, the report “Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy, 2017” outlined the use of various administrative records of the federal and state governments to produce advanced statistics and proposed the activation of inter-departmental coordination and the reorganization of the President's Office of Management and Budget to effectively coordinate government-wide evidence-building efforts.
- The Japanese government has established Evidence Based Policymaking (EBPM) practices, and organized conferences to promote widespread statistical reform that addresses the needs of the people, discuss the era of 'data hegemony' (2017), and recognize various data as public property in order to better address social challenges and drive innovation (Innovation to a Data-driven Society) (Jeong Yong-chan, 2018).
- In the case of Korea, the Ministry of Culture, Sports and Tourism,

the Ministry of Education, the Ministry of Health and Welfare, and the Ministry of Science and ICT established a mid- to long-term plan for the further development of statistics and organized a statistics committee to deliberate on relevant matters for the development of statistics in relevant fields (Jeong Yong-chan et al., 2018).

C. Suggested directionality of governance for big data on women and the family

○ Just as laws regarding statistics are attentive to the organic interdependence of related tasks such as approval of statistical composition, quality control, and establishment of statistical-based policies, big data governance systems must to be designed in a way that enables data collection, linkage, utilization, quality control, and administration specific to big data. The construction of big data governance for women and families will require development in the following areas:

- Increase of dedicated manpower for tasks related to statistics and big data in the Ministry of Gender Equality and Family
- Enactment of regulations on the statistics and data management of women and families
- Designation of a dedicated agency responsible for strengthening the planning and evaluation of women and family statistics
- Institutionalized collaboration between institutions related to statistics and big data of women and families
- Establishment of specific governance for the collection and linkage of statistics and big data for women and families

- The establishment of big data governance for women and the family will require forming a consensus between related agencies, along with changes in work processes such as securing professional manpower, restructuring relevant organization, and promoting awareness for strengthening data-based policies within the Ministry of Gender Equality and Family, all of which will necessitate a phased implementation strategy.

VI. Conclusion and Policy Proposals

1. Proposed Research for Big Data on Women and the Family

- In order to develop beyond a pilot analysis or one-time research, a systematic approach is needed to continuously produce and manage analysis results according to the necessary research topics, and the establishment of big data governance and platforms for women and families will be necessary in order to advance beyond simple analysis to the establishment of predictive and responsive systems. At this time, future big data research on women and the family has been proposed in three major directions: systematization of existing research, discovery of new big data-based research topics, and improvement of big data analysis methods.
- First, most of the existing big data analysis of women and families has been conducted on a trial basis, thus we propose the possibility of expansion by continuing and systematizing further research. In particular, the most frequently used social data analysis can be utilized as an indispensable tool for gathering opinions in the

process of policy planning and enforcement, while the challenge of securing stable data and developing devices for sustained analysis remains. It would also be worth researching how to construct and manage reliable indices for the purposes of managing policy operations. For example, if we create site-specific indicators based on big data analysis of online crimes targeting women and manage them accordingly, we can expect some efficacy in reducing such crimes. In other fields, another important subject for future research could be the development of predictive and responsive systems in the field of safety and welfare. Rather than analyzing particular big data in of itself, the preparation of preventative or post-disaster response systems would be key and in that vein, other research institutes have already proposed establishing systems to prevent disasters or criminal activity.

- Second, exploring new big data women and families will lead to discovery of new and necessary research subjects. Current big data research is inevitably limited by the availability of existing data, therefore new big data and connections therein to women's issues can become new research topics. For example, new topics can be generated from connections to topics such as research on female safety in the workplace, gender equality in the workplace, and female-targeted content by utilizing data on facility buildings, data on the Internet of Things, and administrative data from other departments.
- Third, we propose a study on measures for improvement at the analysis process and methodology level. In order to continuously and effectively utilize big data in the future, it is first necessary to

link existing data in a single nexus and then establish both a system for producing and managing big data as well as a framework for analysis. Secondly, as the use of artificial intelligence spreads, the issue of embedded bias is concurrently emerging, most prominently gender and racial biases. Therefore, research on finding areas where such bias occurs in algorithms and developing methodologies to reduce said bias is an important point of order. Finally, we need to establish measures to increase the utilization of big data. Other research institutes are seeking ways to open up big data analysis results and increase access to information on the terms of prospective users. Research that increases the utilization of big data analysis results by linking social data and secondary data to increase probability of usage and provide information services based on user needs should be conducted in conjunction with big data analysis research.

2. Revision of Regulations for Processing Personal Information

- With the current galvanization of research and industry through big data, issues regarding the protection of personal information and the prevention of personal identification of data have been steadily emerging. The level of regulation regarding personal information is currently relatively high in Korea, yet the ambiguous and non-specific privacy regulations make it difficult to access and apply data due to the limitations of data utilization and distribution.
- In an attempt to address such problems, the European Union (EU) outlined their 2016 General Data Protection Regulations (GDPR) and applies them from 2018 onward, specifically defining the

parameters of personal information, demanding that institutions with personal information evaluate the risks of processing personal information, and taking safety measures such as alias processing and cryptography. Many countries including China and Japan have since introduced similar laws and regulations.

- Korea likewise has a targeted plan for expanding the opening of public data, strengthening and encouraging accessibility through systematic management of public data from such sources as the Public Data Strategy Committee, the Ministry of Public Administration and Security, and the National Statistical Office, but in practice, laws and regulations ensure that there are more data that remain undisclosed or difficult to utilize. Therefore, a more fundamental change in the system is needed, such as future revisions of specific laws and regulations.
- In June 2016, the 'Guidelines for Non-identification of Personal Information' was published to present measures for non-identification of personal information, assessment of appropriateness of re-identification status, and follow-up management for prevention of re-identification, yet these guidelines fell short of having normative legal effects. Amongst laws related to personal information, amendments were proposed to unify the regulatory authority of the Personal Information Protection Act, the Credit Information Utilization and Protection Act, and the so-called Data Protection Act, which were distributed to the Ministry of Science and ICT, the Financial Services Commission, and the Ministry of Public Administration and Security, and passed the plenary session of the National Assembly on December 4, 2019 and

January 9, 2020. The main contents of the amendment are as follows: 1) Introducing the concept of alias information to enable the use of protected data, 2) Streamlining the governance system for personal information protection (governance) by overhauling similar or redundant regulations of related laws and unifying the system of execution, 3) Emphasizing the responsibility of personal information processors in utilizing data, 4) Clarifying the criteria for defining ambiguous 'personal information'. Through these amendments, personal information can be treated anonymously while still being used by relevant institutions and industries, and thus we expect that non-identified data can be used without individual consent in big data analysis.

3. Constructing Governance for Big Data on Women and the Family

- In order to enable the utilization of big data for women and families, policies and processes for managing data availability, usability, integration and security must be adequately addressed, and data governance is required to comply with privacy, security, data quality and management regulations.
- As of yet, the Ministry of Gender Equality and Family does not have such a governance system in place and there is no dedicated agency for this task. The establishment of big data governance for women's families will require the forming of a consensus with related agencies along with changes in work processes such as securing professional manpower, restructuring the organization, and expanding awareness to strengthen data-based policies within the Ministry of Gender Equality and Family. Therefore, the construction of necessary

governance is proposed through the following three-step implementation strategies.

- Step 1: Prepare for the implementation of big data governance
 - Task analysis to enhance expertise in building big data governance
 - Reviewing the feasibility of designating an agency dedicated to statistical planning and evaluation
 - Establishment of regulations on statistics and big data management for women and families
 - Establishing education and public relations measures to form consensus with relevant ministries and agencies
- Step 2: Introduction of the big data governance system
 - Firstly secure statistical professionals and review the feasibility of organizational restructuring for the establishment of big data governance
 - Designate institutions dedicated to statistics and big data on women and family and establish regulation on the management of women and family statistics
 - Conduct education and promotion efforts related to big data governance for internal personnel and related agencies of the Ministry of Gender Equality and Family
 - Launch a pilot project for big data analysis on the theme of women and families
 - Build a big data integrated information system (either through new construction projects or supplementing the role and performance of existing systems)

○ Step 3: Establishing a big data governance system

- Reform that aligns with the establishment of big data governance and the resulting re-organization of tasks.
- Enact regulations on statistics and big data management for women and families
- Internalize education and promotion efforts for the establishment of big data governance for the Ministry of Gender Equality and Family and related agencies
- Expand the big data analysis project for women and families based on the big data analysis pilot project implemented in the introduction period
- Advance the functionality of the big data Integrated Information System

Appendix 1.

National Research Council for Economics, Humanities and Social Studies Cooperative Research Studies

“Using big data on women and family for policy development and improvement (III)”

1. Series in Cooperative Research

	Research Report Title	Institution
19-56-01	Using big data on women and family for policy development and improvement (III)	Korean Women’s Development Institute
19-56-02	Experimental Study and Institutional Initiatives on Gender Discrimination of Big Data Algorithm	Korea University Institute for Research on Labor and Employment
19-56-03	Promoting the Use of Women and Family-related Big Data by Establishing Infrastructure and Data Governance	Korean Information Society Development Institute

2. Research participants

Institution		Chief of Research	Contributing Researchers
Chief Research Institution	Korean Women’s Development Institute	Sung Mi Jung (Associate Research Fellow)	You Kyoung Moon (KWDI Senior Research Fellow) Soo Yeon Lee (KWDI Senior Research Fellow) Ji So Yoon (KWDI Research Fellow) Yeon Gyu Lim (KWDI Research Fellow) Song Yi Park (KWDI Research Fellow) Jae Kyung Lee (KWDI Visiting Researcher)
Contract Research Institution	Korean Information Society Development Institute	Yong-chan Jung (KISDI Research Fellow)	Ji Hyung Shin (KISDI Research Fellow) Se Ran Koh (KISDI Associate Research Fellow) Yoon suk Oh (KISDI Research Fellow) You Kyoung Moon (KWDI Senior Research Fellow)

Institution		Chief of Research	Contributing Researchers
	Korea University Institute for Research on Labor and Employment	Hee Tae Jung (Research Professor, Korea Univ. Institute for Labor Practices)	Sun Up Kim (Korea Univ. Korean Social Research Institute Data Center Manager) Sung Ah Ahn (Prof. Chugye University for the Arts Dept. of Film & Business) Seung Joon Shin (Prof. Korea University Dept. of Statistics)



Korean Women's Development Institute

225 Jinheung-ro, Eunpyeong-gu
(1-363, Bulgwang-dong) Seoul, 03367, Republic of Korea
TEL 02.3156.7000 FAX 02.3156.7007
<http://www.kwdi.re.kr>