Journal of The Korea Institute of Information Security & Cryptology ISSN 1598-3986(Print)
VOL.29, NO.5, Oct. 2019 ISSN 2288-2715(Online)
https://doi.org/10.13089/JKIISC.2019.29.5.1117

개인정보의 활용과 보호를 위한 데이터 거버넌스 성숙도 모형과 종합이행절차에 관한 연구*

이 영 상,[1†] 박 원 환,[2] 신 동 선,[1] 원 유 재[2‡]
[1]데이터스트림즈(주), [2]충남대학교

A Study on Data Governance Maturity Model and Total Process for the Personal Data Use and Protection*

Youngsang Lee,[1†] Wonhwan Park,[2] Dongsun Shin,[1] Yoojae Won[2‡]
[1]Datastreams Corp. [2]Chungnam National Universiry

요 약

최근에 인터넷, 모바일, 사물인터넷 등과 같은 정보기술이 급속히 발전하면서 비즈니스에 필요한 데이터의 수집이 용이해지고, 빅데이터 분석이라는 새로운 방법으로 수집한 데이터를 분석하여 비즈니스에 적절히 활용하고 있다. 이와 같이 데이터의 수집과 분석이 용이해진 반면, 그러한 데이터 속에는 정보주체로부터 직접 제공받지 않은 센서 번호, 기기 번호, IP 주소 등과 같은 식별자 항목이 포함된 개인정보가 관계자도 인지하지 못하는 사이에 수집하여 이용될 수 있다. 만약 이러한 데이터에 대하여 수집에서 폐기까지의 전 과정에서 체계적인 관리가 되지 않을 경우 프라이버시 침해로 인해 「개인정보 보호법」 등 관련 법률 위반이 우려될 뿐만 아니라 데이터 품질 문제도 발생하여 올바르지 못한 의사결정을 할 수도 있다. 따라서 본 연구에서는 기업이 비즈니스에 활용할 목적으로 수집한 데이터 속에 숨겨져 있는 개인정보를 찾아내어 전사적 관점에서 체계적인 관리함으로써, 이를 비즈니스에 효율적으로 활용함과 동시에 보호도 가능하게 하며, 더 나아기 데이터의 품질 확보도 용이하게 하는 새로운 데이터 거버넌스 성숙도 모형(DGMM)과 이를 데이터 거버넌스 프로그램의 도입 또는 개선하고자 하는 현장에 활용하기 위한 종합이행절차를 제안하고, 그 적용 예를 보인다.

ABSTRACT

Recently, IT technology such as internet, mobile, and IOT has rapidly developed, making it easy to collect data necessary for business, and the collected data is analyzed as a new method of big data analysis and used appropriately for business. In this way, data collection and analysis becomes easy. In such data, personal information including an identifier such as a sensor id, a device number, IP address, or the like may be collected. However, if systematic management is not accompanied by collecting and disposing of large-scale data, violation of relevant laws such as "Personal Data Protection Act". Furthermore, data quality problems can also occur and make incorrect decisions. In this paper, we propose a new data governance maturity model(DGMM) that can identify the personal data contained in the data collected by companies, use it appropriately for the business, protect it, and secure quality. And we also propose a over all implementation process for DG Program.

Keywords: Personal Data Use and Protection, Data Governance, Data Governance Maturity Model

Received(06. 11. 2019), Modified(09. 03. 2019), Accepted(10. 13. 2018)

* 본 연구는 "과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업"(IITP-2019 - 2016-0-00304)으로 수행되었습니다.

† 주저자, yslee@datastreams.co.kr
‡ 교신저자, yjwon@cnu.ac.kr(Corresponding author)

I. 서 론

데이터 수집은 입력 단말기 등과 같은 전통적인 방법을 사용하기도 하지만, 최근에는 인터넷, 모바일, 사물인터넷 등 발전된 IT 기술을 이용하여 대량의 데이터를 손쉽게 수집하여 필요한 목적에 부합하도록 분석하여 이용한다.

그러나 이들 데이터는 전사적으로 체계적으로 수집하여 관리되지 않으면, 품질에 관련되는 문제, 공동 활용의 한계, 개인정보의 침해 등과 같은 다양한 위험에 직면할 수 있다. 특히 기업과 공공기관 등(이하 '기업 등')이 수집하여 활용하는 데이터 속에는 정보주체로부터 제공받은 개인정보도 있겠지만, 그 외에도 인터넷, 센서, 사물인터넷 관련 고유번호가 포함된 개인정보를 관계자도 모르는 사이에 포함하여 수집·이용될 수 있다. 이러한 데이터에 포함된 개인정보를 구분·식별하여 체계적으로 관리하지 않고, 또는 정보주체 동의 또는 법률적 근거 등이 부족한 상태에서 비즈니스에 이용하거나 제3자에 제공할 경우, 「개인정보 보호법」(이하 '보호법')등 관련 법령의 위반으로 관계기관으로부터 법률적인 처분과 함께 정보의 주인(정보주체)으로부터 손해배상 피소 등이 우려된다.

이와 같은 데이터 관리 문제를 전사적 관점에서의 해결을 위해 데이터 거버넌스(DG : Data Governance)가 연구되고 있다. 대표적인 연구자로는 Z. Panian[1], G. Thomas[2], S. Soares[3] 등을 들 수 있다. 이들은 각자 DG 모델(프레임워크) 또는 절차를 연구하여 발표하였다. 그러나 이와 같은 모델과 절차에서는 개인정보를 별도로 식별·구분하지 않고, 일반 데이터에 포함하여 취급함으로써 최근에 우리나라와 유럽연합 등에서 강화된 보호법과 제도가 고려되지 않음으로 인하여 데이터 관리 현장에서 적용의 한계가 있다. 따라서 이를 우리나라 개인정보의 활용과 보호 모두에 적합한 새로운 DG 모델과 이를 DG 프로그램의 새로운 도입이나 기존의 것을 개선하는데 필요한 종합이행절차를 제안한다.

본 논문의 2장에서는 DG의 정의, 필요성 등을 소개하고, 3장에서는 DG 프레임워크, 그리고 DG 이행절차에 대한 선행연구를 알아보고, 4장에서는 개인정보의 정의와 이의 식별 방법, 개인정보의 활용과 보호 환경, 그리고 이를 위한 새로운 DG 성숙도 모형과 종합이행절차를 제안하고, 마지막으로 5장에서는 결론과 향후과제를 제시한다.

II. 데이터 거버넌스(DG)

2.1 정의

S. Soares[3]는 데이터를 기업의 자산(asset)으로 취급하여 '기업이 데이터를 전사적 자산으로 활용할 수 있도록 인력(people), 프로세스(process) 및 기술(technology)을 조율하는 활동'으로 정의하고 있으며, 김석수[4]는 DG를 다음과 같이 정의하고 있다.

- 데이터 경영·분배·보호를 향한 프레임워크와 로드-맵
- 데이터 통합과 기업 데이터 경영 프로그램을 지원하기 위한 데이터 수집 전략과 방법
- 산업의 특화된 규정에 정렬하기 위한 접근
- 전반적인 데이터 투명성과 사용을 향한 전략

그리고 J.Ladley[5]는 DG를 '데이터 자산관리에 대한 권한·통제·계획이며, 이를 위한 모니터링이며, 이의 집행과 관련되는 공동의 의사결정'으로 정의하고 있다. 이외에 김석호·이창수[6], 최완일[7], Z.Panian[1], K.Wende[8] 등이 DG를 다양하게 정의하고 있다. 이러한 정의는 산업, 조직, 학술 등 분야별 특성이나 관점, 연구목적 등에 따라 다양하지만, 본 연구에서는 이들을 종합하여 아래와 같이 3가지로 정의한다.

- DG는 기업 등이 자사의 자산인 데이터를 효율적으로 관리·활용하기 위한 원칙이다.
- DG는 기업 등이 조직 내에서 인력, 절차, 기술 및 정책의 조정을 통하여 전사 데이터에서 최적의 가치를 도출하기 위한 방법이다.
- DG는 데이터로부터 야기될 수 있는 데이터의 불일치와 같은 문제로 인하여 상충되는 정책 등을 조정·통제·관리하는 방법이다.

이와 같은 DG를 기업 등이 자사의 비즈니스 이익 극대화를 위해 조직 내부의 업무절차에 적용하여 운영하도록 하는 것을 'DG 프로그램'[3]이라 한다.

2.2 필요성

최근에 기업 등은 관리해야할 데이터가 급증하면서 비즈니스 데이터를 보다 체계적·효율적으로 관리해야

하는 수많은 도전 과제에 직면하고 있다.

이들 과제 중에서 대표적인 몇 가지를 살펴보자. 첫째는 데이터를 자동으로 쉽게 수집할 수 있는 인터넷, 모바일, 사물인터넷 등과 같은 IT 기술의 활용에 따라 관리대상 데이터의 급증 문제를 들 수 있다. 이는 과거의 수작업에 비하여 자동 수집 데이터는 규모면에서 비교가 되지 않을 정도로 대량이기 때문이다. 둘째는 이와 같은 대량 데이터를 식별하고, 분류하고, 저장하고, 통합하는데 소요되는 시간과 비용, 그리고 저장할 스토리지 비용의 증가를 들 수 있다. 셋째는 보호법 등 지켜야 할 법·제도적 규칙이 세계적으로 매년 제·개정되는 20,000여건 이상을 즉시 반영해야 하는 것이다. 특히 다국적 기업인 경우는 통신망을 통한 개인정보의 국외 이전과 같은 문제와 직면할 수 있어 예민한 과제라 할 수 있다. 넷째는 비즈니스 위험을 유발하는 오류 데이터가 대량의 데이터에 포함될 수 있어 일반적인 방법으로는 관리가 불가능하거나 비효율적이어서 적절한 새로운 방법을 모색해야 하는 것이다. 다섯째는 대량의 데이터를 수집하였지만, 이의 활용도가 낮아 비경제적이고, 비능률적이라는 점을 들 수 있다.

따라서 이와 같은 직면 과제의 해결을 위해 기업 등은 수집·보유 중인 데이터의 품질향상, 데이터의 책임성과 보안성 증진, 데이터의 관리에 소요되는 총비용(TCO)의 절감, 오류 데이터에 의한 위험의 감소 등이 가능한 DG에 대하여 많은 관심을 가지고 있으며, 관련 연구도 활발하다.[1-9]

2.3 DG의 범위

기업 등이 DG를 통하여 달성해야 하는 1차적 목표는 앞에서 논의한 당면 과제의 해결일 수 있지만, 실질적인 최종목표는 매출의 증가, 비용의 절감, 규정의 준수(Compliance)를 들 수 있다. 따라서 DG의 범위도 여기에 초점을 두어야 한다.

Fig.1은 비즈니스 관점에서 기업 등이 조직을 운영하는 과정에서 데이터를 주고받거나 공유하는 대상, 관련 규정의 준수 등이 DG의 범위[4]이며, 6개 요소로 되어 있다. 각 요소는 기업 등이 최종목표를 달성하기 위해 필요한 영역이며, 이들 간에는 적절한 시기에 적합한 데이터를 공유하는 메커니즘, 데이터 품질과 일관성 유지 등이 포함된다. 특히 정보보호와 프라이버시 요소는 조직이 세심하게 관리해야 하는 영역이다.

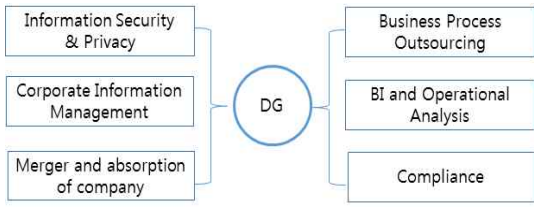


Fig. 1. Coverage of Data Governance[4]

이들 요소 간에 취급하는 데이터의 유형은 일반문서와 전자문서, 그리고 정형 데이터와 비정형 데이터가 있다. DG는 기업 등의 활동에서 생성되는 모든 데이터가 대상이지만, 본 연구에서는 일반 문서 데이터와 비정형 데이터를 제외한 전자문서 데이터와 정형 데이터를 중심으로 연구한다.

III. DG 성숙도 모형과 이행절차

전사적 관점에서 기업 등이 데이터를 체계적으로 관리함으로써 가치 극대화, 일관성 확보 등으로 기업의 경쟁력을 강화하고, 민첩성을 유지하고, 고객의 요구를 능동적으로 충족시키며, 관리비용을 최소화할 수 있다. 즉 기업 등이 수집하여 보유 중인 데이터(이하 '데이터')를 여러 응용 시스템, 비즈니스 절차 등 조직 전체 사용자와 공유하고, 재사용이 가능하기 위해서는 이들 데이터의 체계적인 관리가 필요하다. 따라서 기업 등은 데이터의 사용·개발·관리를 위한 표준, 정책 및 절차 계획을 수립하고, 적합한 관리 인력 조직을 만들어 이행하여야 한다. 이를 위한 방안으로 DG 프로그램을 들 수 있다.

DG 프로그램은 기업 등의 현재 데이터 관리수준을 평가하고, 부족한 내용을 식별하고, 해결과제를 도출하여 개선하는 순으로 진행한다. 이 과정에서 필요한 것 중에서 중요한 것으로 DG 성숙도 모형(DGMM : DG Maturity Model)을 들 수 있다. 이는 기업 등이 DG 프로그램 도입의 준비와 이행에 필요한 분야별 기준이나 방법을 찾는데 도움을 주는 도구이다. 그러므로 DGMM을 이용함으로써 기업 등이 DG로 얻고자 하는 매출의 증가, 비용의 절감, 규정의 준수를 위한 전략적 방법과 기준의 도출이 가능해야 한다.

이러한 관점에서 볼 때, DGMM은 데이터로부터 발생하는 다양하고 복잡한 문제를 식별·평가·해결해야 하는 분야들로 구성된 'DG 프레임워크'이며, DG 성숙

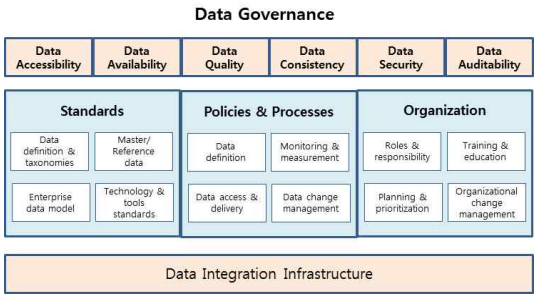


Fig. 2. DG Framework[1]

도 평가와 개선과제 도출 등 실질적인 DG 프로그램을 이행하는 프로세스를 'DG 이행절차'라 한다.

3.1 DG 프레임워크 사례 연구

DG 프레임워크는 적용하고자 하는 환경과 여건에 따라 다양하게 발표되고 있지만, 이중에서 사례로 Z. Panian[1], G.Thomas[2], S. Soares[3]가 제안한 것을 살펴본다.

첫째로 Z. Panian의 DG 프레임워크는 Fig. 2와 같이 전사적 관점에서 데이터는 접근성, 유용성, 품질 등 6가지의 특성을 가져야 하고, 이를 위해 필요한 표준, 정책과 프로세스, 조직, 기술 등 4가지 구성요소를 포함하도록 하고 있다.

둘째로 G.Thomas가 제안한 DGI DG 프레임워크는 10개의 구성 요소를 3개 영역으로 나누어 구성하고 있다. 제1 영역인 '규칙과 업무규칙'과 관련하여 미션과 비전, 목표 데이터 규칙, 결정권, 책임, 통제 등 6개 요소를 중심으로 규칙을 설명하고 있다. 제2 영역인 'DG 조직'과 관련해서는 데이터 이해관계자, 데이터 거버넌스 오피스(DGO), 데이터 스튜어드 등 3개 요소를 이용하여 사람과 조직에 적용하는 '참여 규칙'을 설명하고 있다. 마지막으로 제3 영역인 'DG 프로세스'와 관련해서는 프로세스 요소만을 포함하고 있으며, 데이터를 관리하는데 사용하는 방법을 설명하고 있다. 여기서 프로세스는 표준화와 문서화되어야 하고, 반복이 가능하도록 하고 있다. 그리고 데이터 관리, 개인정보 보호, 정보보안 등에 대한 규정 준수를 지원할 수 있도록 하고 있다.

셋째로 S. Soares가 제안한 'IBM의 DGMM'이라는 프레임워크다. 이는 Fig. 3과 같이 4개 영역에서 11개 요소로 설계되어 있다. 최종 목표인 'Outcomes'

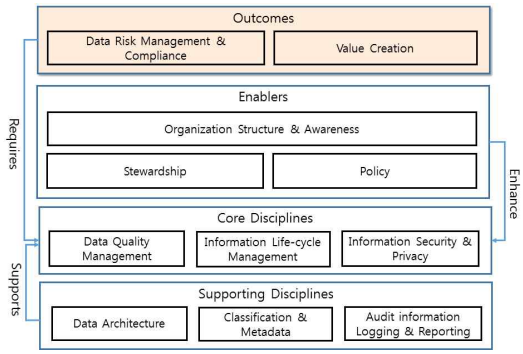


Fig. 3. DGMM of IBM[3]

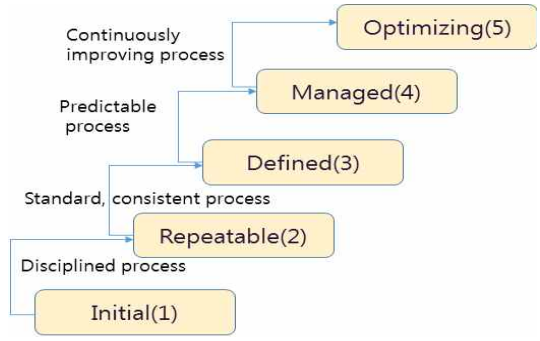


Fig. 4. Software Capability Maturity Model[10]

영역에 'Data Risk Management & Compliance'와 'Value Creation' 요소를, 'Enablers' 에는 'Organization Structure & Awareness' 등 요소를, 'Core Disciplines' 영역에는 'Data Quality Management' 등을, 'Supporting Disciplines' 영역에는 'Data Architecture' 등을 포함하고 있다.

3.2 DG 성숙도 평가의 등급 기준

DG의 성숙도를 평가하는데 사용하는 등급 기준은 알려진 모델이 없다. 그러나 소프트웨어 개발의 참여업체를 객관적으로 평가하는 기준을 카네기멜론 대학과 소프트웨어 공학 연구소(SEI : Software Engineering Institute)가 제안한 CMM (Capability Maturity Model)[10]과 이를 확장한 CMMI[11]가 알려져 있으며, 이 모델을 다양한 분야에서 응용하고 있다.

CMM은 Fig. 4와 같이 개발 수준을 다섯 단계, 즉 초기(rnitial)·반복(repeatable)·정의(refined)·관리(managed)·최적화(optimizing) 단계로 나누고 있

다. 이 다섯 단계는 소프트웨어 기술을 사용하는 조직의 역량을 측정하는 성숙도 수준이다. 이 모델을 DG 성숙도 수준 평가에도 응용이 가능하며, 그 사례[3]도 있다.

3.3 DG 프로그램의 이행절차

기업 등은 앞에서 논의한 DG 프레임워크를 이용하여 자사의 현재 DG 성숙도 평가를 거쳐 적절한 절차에 따라 DG 프로그램을 진행할 수 있다. 이때의 일반적인 절차는 DG 프레임워크의 각 요소별로 현재 DG 상태와 미래의 희망 목표를 파악하고, 두 상태 간의 격차(gap)에 따른 문제점과 그의 해결과제를 도출한 후, 각 요소별 과제의 개선계획을 마련하여 이행하는 순으로 진행된다.

앞의 세 가지 프레임워크 중에서 S. Soares도 자신의 연구에서 정한 프레임워크와 적용 절차를 함께 제안하고 있다. 그는 DG를 현업에 적용하는 절차를 비즈니스 문제의 인식 및 정의(1단계), 중역 스폰서십 확보(2단계), 성숙도 평가(3단계) 등을 거쳐 최종으로 결과측정(14단계)까지 총 14단계로 정의하고, 각 단계별로 이행해야 하는 내용도 함께 제시하고 있다.

3.4 DG 프레임워크와 개인정보의 활용과 보호

앞에서 살펴본 세 가지 DG 프레임워크에 포함되어 있는 'Security' 또는 'Information Security and Privacy' 요소가 있지만, 이는 정보와 개인정보의 안전한 보호를 위해 필요한 정보보호 기술의 적용에 적합하도록 고안되어 있다. 이는 Table 1에 있는 OECD 개인정보보호 8원칙(이하 '8원칙')[12]의 '(5) Security Safeguards Principle' 원칙에는 직접적으로 관련되지만, 그 외의 원칙들과는 직접적인 관련이 없다.

이 8원칙 모두는 개인정보를 수집하는 단계에서부터 파기까지의 모든 과정에서 체계적으로 다루어져야 한다. 실제로 유럽연합(EU)의 GDPR(General Data Protection Regulation)[13]이나 우리나라의 보호법은 이 8원칙에 준하여 관련 규정을 두고 있다. 하지만, 앞에서 살펴본 DG 프레임워크에서 이 8원칙 모두를 적용하는 것은 쉽지 않다. 예를 들어, '(1) Collection Limitation Principle(수집제한의 원칙)'은 개인정보를 수집할 때, 비즈니스에서 반드시 필요한 개인정보 항목만을 적법하고 공정한 수단에 의하여 수집해야 하는 원칙이다. 또 다른 예로 '(2) Data Quality Principle(데이터 정확성의 원칙)'은 수집하는 개인정보가 사용목적과 범위에 부합해야 하고, 정확하고 안전하게 수집·갱신·유지되어야 한다는 원칙이다. 이들 원칙 모두를 살펴볼 때, 앞의 DG 프레임워크들은 8원칙 중에서 '개인정보의 안전한 보호'에 해당하는 다섯째 원칙과는 관련이 있지만, 그 외의 원칙들은 고려되지 않고 있다.

Table. 1. OECD the Privacy 8 Principles

(1) Collection Limitation Principle There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means and, where appropriate, with the knowledge or consent of the data subject.
(2) Data Quality Principle Personal data should be relevant to the purposes for which they are to be used, and, to the extent necessary for those purposes, should be accurate, complete and kept up-to-date.
(3) Purpose Specification Principle
(4) Use Limitation Principle
(5) Security Safeguards Principle Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorised access, destruction, use, modification or disclosure of data.
(6) Openness Principle
(7) Individual Participation Principle
(8) Accountability Principle

따라서 기업 등이 DG 프로그램을 도입할 때, 이 8원칙들 모두가 적용이 가능하도록 개인정보의 수집에서부터 파기까지 전 과정에서 활용과 보호를 함께 고려한 새로운 DGMM이 필요하다.

IV. 개인정보의 활용과 보호를 위한 DGMM

개인정보를 수집하여 이용하지 않는 국내의 기업 등은 거의 없으므로 모든 기업 등은 보호법이나 「정보통신망 이용촉진 및 정보보호에 관한 법률」(이하 '정보통신망법')과 같은 법률의 적용을 받는다. 또한 EU 내의 국가에서 사업을 하는 기업 등은 GDPR의 적용도 받는다.

이들 법률의 규정을 위반하면 형사적·민사적 처분을 받으며, 특히 GDPR의 경우는 세계 각국의 어떠한 기

업이든 전 세계 매출의 최대 4%까지 과태료 처분을 받을 수 있다. 여기서 위반이라 함은 개인정보의 보호를 소홀히 하여 외부로 개인정보를 유출한 경우만 해당되는 것이 아니며, 개인정보의 수집에서부터 파기까지의 전 과정에서 관련 규정을 위반한 경우도 포함한다.

하지만, 이와 같은 DG 관점에서 데이터 관리 문제도 있지만, 그 보다 더 근본적인 문제는 기업 등이 수집·활용하는 데이터 내에 어떠한 것이 개인정보에 해당하는지를 명확히 알 수 없거나, 알고 있다고 하더라도 각 비즈니스별로 별도로 수집·이용하던 개인정보를 빅데이터 분석 등을 위해 전사적 관점에서 다른 데이터와 통합 또는 결합하여 이용할 경우에 어떤 것이 개인정보에 해당하는지 불분명해진다. 이러한 경우에 어떠한 것이 개인정보에 해당하는 지를 명확히 식별하여 DG 관점에서 규제준수 등 관련 DG 관련 요소와 함께 관리되어야 한다.

따라서 본 논문에서는 개인정보의 식별 방법, 그리고 개인정보의 수집에서 파기까지 전 과정에서 활용과 보호를 함께 고려하고, 8원칙을 준용하는 보호법과 GDPR 등에 적합한 새로운 DGMM과 이를 현장에서 활용을 위한 종합이행절차를 연구한다.

4.1 개인정보 정의와 이의 판단방법

개인정보는 보호법 제2조 제1호에서 "살아 있는 개인에 관한 정보로서 성명, 주민등록번호 및 영상 등을 통해 개인을 알아볼 수 있는 정보(해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 것을 포함한다)를 말한다."라고 정의하고 있다. 또한, 동조 제4호에서 개인정보파일(PDF : Personal Data File)을 "개인정보를 쉽게 검색할 수 있도록 일정한 규칙에 따라 체계적으로 배열하거나 구성한 개인정보의 집합물(集合物)을 말한다."라고 정의하고 있다.

이 개인정보와 PDF의 정의에 따라 기업 등이 보유 또는 보유 예정 데이터가 개인정보에 해당하는지를 판단해야 하지만, 단순히 데이터 외형만으로 판단하는 것은 쉽지 않다. 실제로 데이터 관리 현장에서 그러한 어려움을 겪고 있다. 이를 위해 가상(Virtual)의 파일(DF : Data File)을 (E.1)과 같이 정의한다.

$$DF = \{A_1, A_2, A_3, \cdots, A_n\} \quad \text{(E.1)}$$

여기서 DF는 n개 항목(A_i)으로 구성된 집합이며, 기업 등이 보유(또는 보유 예정) 데이터 항목들(Items)이다. A_i는 실제 값이 저장되는 변수(Variable) 또는 애트리뷰터(Attribute)이다.

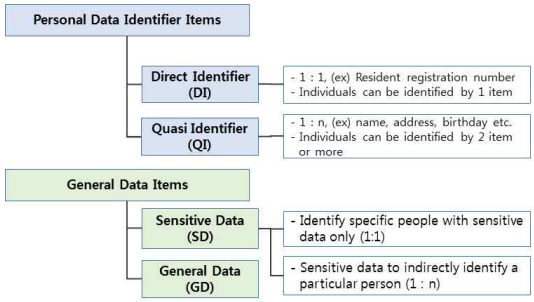


Fig. 5. Properties of PDF items[14,15]

DF가 PDF에 해당하는지는 A_i(i=1, … , n)에 실제로 저장되는 값(실제 값)에 따라 판단하여야 한다. 그 이유는 A_i에 '이름', '전화번호', '주민등록번호' 등과 같은 항목을 포함하고 있지 않으므로 PDF에 해당하지 않는다고 판단할 수 있지만, 실제로 MAC(Media Access Control) 주소와 같은 특정 개인의 식별이 가능한 항목, 즉 직접 식별자(DI)가 포함되어 있을 수도 있기 때문이다. 따라서 DF가 PDF에 해당하는지는 DF의 각 항목의 실제 값을 이용하여 Fig. 5와 Table 2에 따라 판단할 수 있다.

Fig. 5의 식별자(Personal Data Identifier) 항목은 직접 식별자(DI : Direct Identifier)와 준(간접) 식별자(QI : Quasi Identifier)로, 일반 데이터 항목(General Data Items)은 민감 정보(SD : Sensitive Data)와 일반 데이터(GD : General Data)로 구분된다. DI는 실제 값 자체만으로 특정 개인의 식별이 가능한 주민등록번호, 전화번호, MAC 주소, IP 주소 등을 말한다. QI는 실제 값만으로 특정 개인을 식별할 수 없지만, 두 개 이상의 값이 결합하여 특정 개인의 식별이 가능한 항목으로 이름(동명이인이 있을 경우), 주소, 생년월일 등을 말한다. 민감 정보(SD : Sensitive Data)는 보호법제23조에서 "사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활 등의 정보"라고 정의하고 있다. 하지만, 이는 SD에 해당하는 단독 항목만의 경우는 일반정보에 해당하며, DI(또는 두 개 이상 QI)과 결합한 경우에 PDF에 해당함과 동시에 SD와 개인정보에도 해당한다.

그러나 DI와 QI에 해당하는 항목들을 법령에서 정하고 있지 않으며, 가이드라인[14]에 일부 사례만 있

다. 그러므로 이들 식별자는 기업 등이 스스로 판단해야 하는 어려움이 있다. 따라서 법령에서 구체적으로 정할 필요가 있으며, 이를 위한 연구는 별도로 요구된다.

Table 2를 참고하여 DF가 PDF에 해당하는지 여부를 판단할 수 있다. 여기서 '+'는 항목들의 결합을, 'Other data item(s)'는 해당 항목을 제외한 다른 항목들을 의미한다. Table 2와 Fig. 5는 보호법의 PDF 정의와 관련 연구[14,15]를 참고하여 본 연구에서 데이터 관리자가 활용이 가능하도록 알고리즘 형식으로 변형한 것이다. 만약 이를 이용하여 가상(임의)의 DF가 PDF로 판단되면, 그 파일에 포함되어 있는 모든 레코드는 개인정보에 해당하며, 반대일 경우는 일반데이터에 해당한다. 그러나 일반데이터파일에 포함되어 있는 모든 레코드가 일반정보라고 볼 수 없다. 그 이유는 독특한 값의 조합으로 만들어진 특수한 레코드가 아주 드물게 존재할 수 있기 때문이다. 이러한 경우를 위하여 k-익명성[14] 등의 연구가 있지만, 본 연구의 범위를 벗어난다.

또 다른 경우로 개인정보의 수집이나 생성 과정에서 항목의 증감이 발생할 경우에도 해당 DF가 PDF에 해당여부를 판단해야 한다. 이때의 가상의 파일 DF1을 (E.2)로 정의한다.

$$DF1=\{A_1,\cdots,A_{n1},B_1,\cdots,B_{n2},C_1,\cdots,C_{n3}\} \quad \text{---} \ (E.2)$$

위 (E.2)의 DF1은 업무나 서비스를 제공받기 위해 정보주체가 기업 등에 제공한 정보의 항목(A_1, … , A_{n1}), 이를 기초로 기업 등이 업무나 서비스를 제공하는 과정에서 생성되는 정보의 항목(B_1, … ,B_{n2}), 그리고 위 두 항목들 정보를 기초로 빅데이터 분석 등으로 가공·재생산한 정보의 항목(C_1, … ,C_{n3})으로 구성되어 있다. 위 DF1의 경우도 DF와 동일한 방법으로 PDF와 개인정보 여부를 판단한다.

Table 2. Criteria for Personal Data Files

(1) DI only → Personal Data File - ex : 'Resident registration number', 'IP', 'Phone number', etc.
(2) DI + Other data item(s) → Personal Data File - ex : 'Resident registration number' + 'name'
(3) QI only → General Data File - ex : 'name', 'address(dong)', 'birthday', etc.
(4) QI + Other data item(s)→ Personal(General) Data File - ex : 'name'+'address(dong)'+'salary'+ 'height' + 'weigh' → Personal Data File
(5) SD(s) only → General Data File - ex 'Political party affiliation', 'Joining labor union, etc (6) SD(s)+GD(s) → General Data File - ex : 'Joining labor union' + 'height' +'weigh' (7) GD(s) → General Data File

이러한 예는 스마트폰 개통을 위해 가입자(A)가 신청서에 이름(A), 주소(xxx동 zzz번지), 주민등록번호(xxxxxx - yyyyyyy) 등을 기입, 통신사에 제출하여 스마트폰을 개통·사용하면, 통신사는 과금 등을 위해 통화기록, 전화요금, 요금납부 여부 등의 정보를 수집하고, 필요시에는 이들 정보를 분석·가공하여 재생산한 정보(예: 고객의 특성 등)를 재생산하여 업무에 활용할 수 있다. 이와 같은 모든 항목을 포함하는 파일(DF1)은 Table 2의 (1), (2)에 따라 PDF에 해당하고, PDG 파일의 모든 레코드는 개인정보에 해당한다. 그리고 (E.1)이나 (E.2)의 일부 항목으로 구성된 파일(Subset)의 경우도 Table 2의 기준에 따라 PDF에 해당하는지 여부는 다시 판단하여야 한다.

기업 등은 자신들이 보유·관리 중인 데이터베이스가 PDF에 해당여부를 판단하고자 한다면, 모든 데이터를 대상으로 프로파일링 방법 등으로 DF(또는 DF1)를 생성한 후, 위의 방법과 기준에 따라 PDF와 개인정보 여부를 판단하여야 한다. 그 이유는 일반적으로 데이터베이스를 설계할 때, 시스템의 성능 등을 고려하여 하나의 데이터베이스 스키마를 여러 개의 테이블로 분할하는 경우, 또는 여러 업무에서 동일한 내용의 테이블을 중복 사용하는 경우가 있기 때문이다. 이와 같이 전사적 관점에서 PDF 여부를 판단해야 하는 또 다른 이유는 보호법 제2조 제5호에서 '개인정보처리자'의 범위가 "공공기관, 법인, 단체 및 기업 등"이라고 정하고 있어 해당 기업 등에서 수집하는 모든 개인정보를 종합하여 판단해야하기 때문이다. 관련 연구로는 황수하·김정덕[16]을 들 수 있다.

4.2 개인정보 활용과 보호 환경

보호법이나 GDPR은 개인정보의 수명주기, 즉 개인정보의 수집, 저장·관리, 이용·제공, 파기의 전 과정[12]을 규율한다. 이에 따라 DGMM의 요소들 중에는 개인정보의 수집·이용과 관련되는 요소와 보호 관련 요소가 모두가 포함되어야 하고, 그 수준의 평가도 가능해야 한다.

이를 위해 개인정보 수명주기의 각 단계를 간단히

살펴보면, 첫째로 '수집'은 처음으로 정보주체 등으로부터 개인정보를 수집하는 단계이다. 이 단계에서는 우선 앞 절 논의한 DF(또는 DF1)를 정의하고, 각 항목의 실제 값에 따라 속성정보, 즉 직접 식별자, 간접 식별자, 민감 정보 등을 구분하여 Table 2에 따라 PDF 여부를 판단한다. 이 DF의 부분집합을 사용할 경우도 동일하다. 각 항목의 속성 정보(DI, QI 등)는 메타 데이터에 해당한다. 둘째로 '저장·관리'는 이미 수집·활용한 개인정보는 다음의 또 다른 활용을 대비하거나 법적 보관의 의무이행을 위해 데이터베이스 등의 형태로 저장·관리하는 단계이다. 이 단계에서는 해당 파일에 대한 접근권한 관리, 암호화, 접속기록관리 등 안전성 확보[17]에 필요한 조치를 요구하고 있어 다양한 정보보호 기술이 요구된다. 셋째로 '이용·제공'은 저장·관리중인 개인정보를 자신의 비즈니스에 이용하거나 제3자의 요청에 대비하는 단계로서 이용·제공의 근거, 즉 동의 또는 법적 근거가 명확할 때 이용이나 제3자 제공이 가능하다. 이 때 제공의 근거가 부족할 경우에는 개인정보의 비식별 조치[14]를 거쳐 제공할 수 있다. 넷째, '파기'는 수집한 개인정보가 당초 목적을 달성한 경우에는 법적으로 보관의무가 있는 경우를 제외하고는 해당 개인정보를 삭제·파기해야 하는 단계이다. 파기의 방법은 데이터베이스 내에 존재하는 각 레코드별로 삭제하는 방법과 데이터베이스 전체를 삭제하는 방법이 있다. 이를 위해 매일 또는 정기적으로 PDF에서 파기 대상을 레코드별로 식별해야 한다.

이와 같은 내용을 감안하면, 기업 등은 수집·보관·이용하거나 그러할 모든 데이터(Fig. 6)에서 개인정보를 식별하여 수명주기에 따라 활용과 보호를 위한 활동을 하여야 한다. 그러한 활동의 일환으로 기업 등이 DG

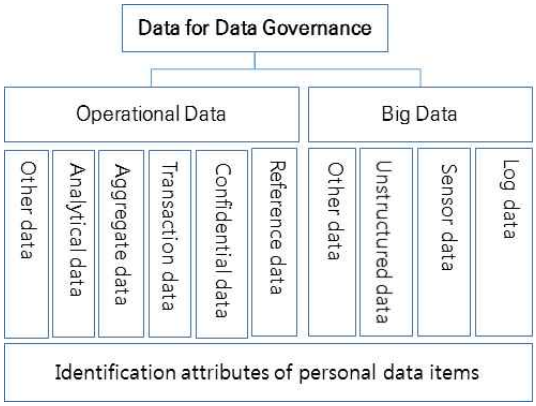


Fig. 6. Data for Data Governance

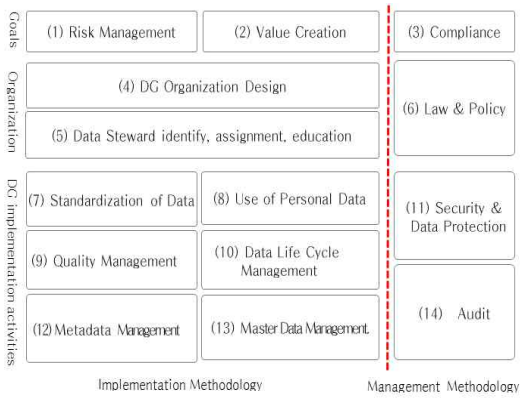


Fig. 7. New DGMM for use and protection of personal data

프로그램의 도입을 검토할 수 있으며, 이때 그러한 활동의 수준이 현재 어느 정도에 해당하는지, 그리고 부족한 부분이 무엇인지를 알아보는 성숙도 평가가 필요하다. 이러한 경우를 대비하여 DGMM에 개인정보의 활용과 보호를 위한 요소가 포함되어야 한다. 기업 등은 이 DGMM을 이용하여 스스로 또는 전문기업을 통하여 보유한 Fig. 6과 같은 모든 데이터 속에 포함되어 있는 개인정보 식별자 항목을 찾는 절차와 방법 등의 존재 여부를 알아볼 수 있을 뿐만 아니라 부족한 부분이나 개선사항 등을 명확하게 파악할 수 있다. 물론 이러한 DGMM이 없어도 관련 업무의 수행은 가능하지만, 객관적·체계적·일관성 등을 기대하기 어렵다.

4.3 개인정보의 활용과 보호를 위한 DGMM 제안

개인정보의 활용과 보호를 위하여 Fig.7과 같은 DGMM(이하 '제안 DGMM')]을 제안한다. 이는 기업 등에서 개인정보의 활용과 보호에 중점을 둔 DG의 성숙도를 평가에 활용할 수 있는 새로운 프레임워크이다. 이는 기존의 DG프레임워크의 요소에는 없는 'Use of Personal Data(개인정보의 활용)'와 'Security and Data Protection(정보보호와 개인정보의 보호)'를 분리하고, 'Standardization of Data(데이터 표준화)'도 별도로 두어 데이터의 공동 활용이 용이하도록 하였다. 그리고 위험에는 비즈니스, 문화, 규제 위험[3]이 있지만, 이중에서 규제 위험을 'Compliance (규정준수)'로 분리하고, 비즈니스와 문화적 위험은 'Risk Management(위험관리)'에 포함되도록 하였다. 그 결과 제안 DGMM은 총 14개 요소가 되었다.

제안 DGMM 형태에서 가로 방향으로 (1)~(3)은

'Goals(DG목표)'를, (4)~(6)는 'Organization(DG 조직)', (7)~(14)은 'Implementation Activities (이행활동)' 분야로 구분하였다. 세로 방향으로 점선의 좌측에는 'Implementation methods(이행방법)'을, 우측에는 'Management methods(관리방법)'으로 구성하여 각 요소별 역할도 구분하였다.

이 제안 DGMM의 각 요소별 내용은 아래와 같으며, 이를 기초하여 관련 정보를 수집·분석하여 성숙도 평가에 활용할 수 있다.

- (Risk Management) 데이터로부터 유발되는 비즈니스와 문화적 위험을 식별·평가·정량화·회피·수용·완화·양도하는 방법과 달성할 목표 등을 말한다.
- (Value Creation) 데이터 자산의 가치를 향상시키고, 평가하고, 계량화한다. 이는 비즈니스 이익 극대화에 기여한다.
- (Compliance) 데이터와 관련되는 모든 법령 등의 규정을 준수하기 위한 절차와 방법, 이를 통해 취할 수 있는 목표를 정의한다. 특히 개인정보의 활용과 보호 관련 규정의 준수를 위한 정책과 방법도 포함한다.
- (DG Organization Design) 비즈니스 조직과 IT 조직 간의 책임과 역할을 명확히 정의한 DG 조직을 설계하여 운영하는 절차와 방법이다.
- (Data Steward identify, assignment) 데이터를 분야별로 분류하고, 각 분야별 데이터의 가치 향상, 위험 완화, 조직 통제 등을 수행할 담당자(스튜어드)를 지정하고, 해당 직무 또는 행동기준을 정한다. 그리고 필요할 경우는 관련 교육도 실시한다.
- (Law & Policy) Compliance(규정 준수)에서 정의한 정책이나 기준을 관리·운영하는 조직의 정의와 함께 구성원의 행동기준을 정의한다. 물론 보호법 상의 조직과 인력(예: 개인정보보호책임자 등)도 포함·정의하고, DG 관점에서 역할과 기능, 그리고 활동 기준의 정의도 포함한다.
- (Standardization of Data) 기업 등에서 사용하고 있는 용어(비즈니스, 기술), 데이터 관련 기준 등을 전사적으로 공동 활용이 용이하도록 표준화 작업을 하고, 그 결과를 메타 데이터 저장소(repository)에 저장·관리한다.
- (Use of Personal Data) Fig. 6과 같은 DG 대상 데이터 전체에 대하여 DF(또는 DF1)를 생성하고, 각 항목이 어떤 식별자에 해당하는지를 판단한다. 만약 개인정보 식별자(DI 또는 QI)로 판단된 항목이 있으며 Table 2에 따라 PDF 여부를 판단하고, 그의 부분집합이나 결합(Join) 가능한 다른 테이블의 레코드는 개인정보로 정의한다. 또한 식별된 개인정보는 앞 절에서 논의한 내용으로 활용을 위한 절차와 방법을 정의한다.
- (Data Quality Management) 데이터의 수집이나 생산, 데이터 정제, 테스트 등 데이터 품질 확보 활동에 해당하는 품질측정·품질향상·품질 인증 등을 위한 요소이다. 이는 OECD 8원칙 중 'Data Quality Principle'과 보호법의 제3조 제3항과도 관련된다.
- (Data Life-cycle Management) 데이터의 수집·이용·저장·삭제에 대한 정책을 말하며, 개인정보의 수명주기와 관련되는 정책과 절차도 포함한다. Fig. 8의 4.2를 참고한다.
- (Security & Privacy) 규정 준수(compliance)에서 정한 정책 등을 이행하는 실질적인 방법이며, 데이터를 보호하기 위한 세부정책과 사례, 통제 규칙 등을 포함한다. 여기에 개인정보의 보호는 수집·이용 중이거나 보관 중인 개인정보에 대하여 유출이나 노출 등에 대비하는 보호기술을 정의·도입·적용도 포함한다.
- (Metadata Management) 비즈니스와 IT 기술의 용어 정의, 분류, 데이터 모델 자료, DF(또는 DF1)의 항목속성 정보 등과 같은 메타 데이터를 정의하고, 이들을 체계적·효율적 관리를 통해 보유 데이터의 이용을 지원한다.
- (Master Data Management) 마스터 데이터는 전사적으로 공동 활용하는 기준 또는 핵심 데이터이다. 이와 관련한 자세한 내용은 Fig. 8의 4.1을 참고한다.
- (Audit) 정한 평가지표(KPIs: Key Performance Indicators)를 측정하여 DG 프로그램이 제대로 동작하도록 감독한다. 여기서 측정한 결과는 최상위 조직인(가칭) 'DG 위원회'에 보고· 공유하는 것이 중요하다. 물론 보호법 제29조(안전성조치의무), 「개인정보의 안전성 확보조치 기준」 고시 등 규정의 준수 여부도 측정·감사한다.

4.4 DG 성숙도 평가

DG 성숙도 평가는 기업 등이 수집·보유·이용하는 모든 데이터를 대상으로 DG 정책과 방법이 얼마만큼 조직 전체에 반영되고 있는지, 그리고 부족한 것이 무엇인지를 알아보는 것을 말한다. 이는 DG와 관련되는 모든 분야에 대하여 객관적이고 체계적이어야 하므로 제안

DGMM을 활용하여 평가하는 것이 좋은 방법이다.

그 방법은 제안 DGMM의 14개 요소별로 데이터 관리 관련 정보를 수집·분석하여 성숙도 수준을 평가한다. 이때 수집하는 정보는 세 가지 종류가 있다, 첫째는 관계 임직원과의 인터뷰 정보이고, 둘째는 해당 기업 내 업무 관계자에게 요청하여 제공받은 정보, 그리고 마지막으로 진단도구로 자동으로 획득한 정보가 있다. 여기서 수집한 모든 정보는 제안 DGMM의 각 요소별로 분석하여 평가하며, 그 결과는 Table 3에 따른다. 평가 결과는 현재의 수준과 미래의 목표 수준, 그리고 그들 간의 격차(gap)가 된다. 이는 DG 프로그램의 이행 과제의 도출 또는 후속 작업의 기초자료로 활용한다.

Table 3은 CMM의 5단계(Fig. 4)를 DG 성숙도 평가에 적용할 수 있도록 변형한 것으로 'Initial'은 DG 프로그램을 도입하지 않은 기업 등이며, 'Optimized'는 최상위 수준의 기업 등이 해당한다.

Table 3. DG Maturity Level

(1) Initial · In the absence of application of DG policy etc. · Each element of the framework does not apply.
(2) Repeatable · A state that is repeatedly applied only to work procedures that succeeded in DG · Some of the detailed procedures of each element of the framework are applied
(3) Defined · All DG programs are defined and applied in the company. · Frameworks Applying each element at the entire company.
(4) Managed · Management procedures such as performance measurement, analysis, and improvement are applied to all DG procedures.
(5) Optimized · Continually improve and optimize DG procedures.

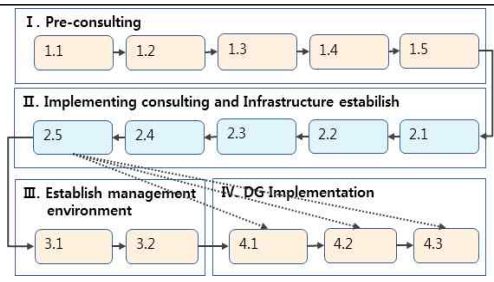
4.5 DG 프로그램의 종합이행절차

이번 절에서는 제안 DGMM의 활용 시기와 방법 등에 대하여 알아본다. 기업 등이 새로이 IT기술을 도입하거나 기존 업무를 개선하고자 할 경우의 일반적인 절차는 먼저 현황 분석을 통해 문제점 파악, 개선과제 도출, 과제별 이행계획 수립 등의 순으로 진행한다. DG 프로그램의 경우도 이와 유사한 Fig. 8의 "DG 프로그램의 종합이행절차(Overall Implementation Process for DG Program)"를 제안한다.

이는 DG 프로그램의 도입이나 개선과 관련하여 총 4개의 영역, 즉 'Ⅰ. Pre-consulting(사전컨설팅)', 'Ⅱ. Implementation consulting and Infrastructure establish(이행 컨설팅과 기반구축)', 'Ⅲ. Establish management environment(관리 환경 구축)', 그리고 'Ⅳ. DG Implementation(구현)' 으로 되어 있다.

첫째 영역인 'Ⅰ.사전컨설팅'은 DG 프로그램을 새로이 도입하거나 기존의 것을 개선을 하고자 하는 경우에 진행하며, 5개의 절차로 되어 있다.

- (1.1 Recognizing data issues for business) 기업 등의 임원이나 중간 간부 등과 면접을 통해 DG 프로그램의 필요성 등 기초적인 비즈니스 데이터 관련 과제를 도출하고 정의한다.
- (1.2 Assessment of DG maturity level) 제안



〈description〉
1.1 Recognizing data issues for business
1.2 Assessment of DG maturity level
1.3 Estimation of data value
1.4 Creating a roadmap for the DG Program
1.5 Approval of the plan of the DG program
2.1 Design of DG personnel organization
2.2 Collection and standardization of terms, and build a metadata repository
2.3 Understanding enterprise data
2.4 Identification of identifiers for the use of personal data
2.5 Metadata collection and management
3.1 Define KPIs
3.2 Security & privacy rule management
4.1 Master data management
4.2 Data life-cycle management
4.3 Data quality management

Fig. 8. Overall Implementation Process for DG Program

DGMM의 각 요소별로 작성한 설문서를 이용하여 비즈니스와 IT 담당자를 대상으로 설문한다. 그리고 이와는 별도로 데이터 관련 현황 자료를 제공받고, 진단도구를 이용하여 시스템에서 관련 정보를 수집하여 설문 내용, 제공받은 자료에 대한 진위 여부를 판단한다. 이들 정보를 종합·분석하여 DG의 성숙도 수준을 평가한다. 이때 수집하는 정보의 범위와 규모는 해당 기업 등의 여건에 따라 가감할 수 있지만, 기본적으로 필요한 내용은 사전에 정의하여 포함시켜야 한다. 이 절차의 평가 결과는 제안 DGMM의 각 요소별로 Table 3을 이용하여 현재 수준과 미래의 목표 수준으로 표현된 내용이며, 그 이유도 명확해야 한다.

- (1.3 Estimation of data value) 성숙도 평가 결과를 참고하여 현재 데이터 가치와 문제점을 개선한 후의 미래 데이터 가치를 추정하여 산정한다. 여기서 산출된 가치는 경영진의 DG의 이해를 구하는데 많은 도움이 되도록 정성적, 정량적으로 명확하고 투명하게 산출하여야 한다.
- (1.4 Creating a Roadmap for the DG Program) 앞 절차에서 도출한 다양한 문제점과 성숙도 평가 결과를 종합하여 개선할 과제를 도출하고, 각 과제별로 개선방안, 일정계획, 소요예산 등을 작성한다.
- (1.5 Approval of the plan of the DG program) 선행 절차의 결과를 기초로 현황 및 문제점, 과제별 개선방안, 추진일정, 소요예산, 데이터 가치의 변화를 포함한 기대효과 등을 종합·정리한 '(가칭)DG 프로그램 도입계획(안)'을 작성하여 최고의사결정자로부터 승인을 득한다.

둘째 영역인 'Ⅱ. 이행 컨설팅과 기반구축'에서는 승인을 득한 "DG 프로그램 도입계획(안)"에 대한 세부이행계획의 수립과 이행을 준비한다. 이 영역에는 총 5개 절차가 있다.

- (2.1 Design of DG personnel organization) DG 프로그램의 이행에 필요한 인력조직을 설계하여 구성하는 단계이다. DG 관련 조직의 예는 최고 의사결정 기구인 'DG 위원회', 중간간부가 참여하는 '중견관리자 지원반', 데이터 유형별로 모든 권한을 갖는 '스튜어드(steward)' 등이 있다. 그리고 DG 프로그램과는 직접 관련은 없지만, 보호법에서 정하고 있는 개인정보 보호책임자(CPO), 개인정보취급자 등을 포함하도록 설계한다.
- (2.2 Collection and standardization of terms, and build a metadata repository) 기업 등이 사용하는 비즈니스 용어와 IT 기술 용어를 수집하고, 이를 표준화하여 공유 가능한 용어사전을 만드는 절차이다. 그 결과는 메타데이터 저장소에 저장하여 관리하면서 전사적으로 공유한다.
- (2.3 Understanding enterprise data) 기업 등의 수집·이용하는 모든 데이터에 대한 목록을 수집·분석하고, 관련 정보시스템도 분석하여 마스터 데이터 설계의 기초자료를 수집하는 단계이다. 이 절차의 결과는 PDF와 개인정보의 식별과 데이터 표준화 작업에 이용될 수 있다.
- (2.4 Identification of identifiers for the use of personal information) 앞에서 알아본 바와 같이 기업 등에서 수집·이용하는 데이터의 상당부분이 보호법 등 관련 법률의 정의에 따른 개인정보에 해당할 수 있다. 따라서 전사적으로 관련 데이터 항목을 연계하여 연계성을 알아보는 프로파일링 기술을 이용하여 DF(또는 DF1)을 생성한 후, 각 항목의 속성을 파악하여 PDF 해당여부를 판단한다. 만약 해당 DF가 PDF에 해당하면, 수집의 근거가 관련 법령 규정(동의, 법령상의 의무 등)의 적절성도 판단한다. 만약 제3자에 제공하는 데이터가 있다면, 그 데이터도 PDF 해당 여부를 판단한다. 만약 PDF에 해당한다면, 개인정보의 비식별 조치 등의 방법으로 변환하여 제3자에 제공하도록 한다. 이 과정에서 식별하거나 분석한 결과는 메타데이터로 관리하면서 전사적으로 공유하도록 한다.
- (2.5 Metadata collection and management) 비즈니스 용어와 기술 용어를 수집하여 표준화한 결과인 메타 데이터, 그리고 PDF 항목의 속성 등과 같은 메타데이터를 수집하고, 이를 체계적으로 관리·공유할 수 있도록 한다. 이때 수집한 메타데이터는 이후의 마스터 데이터 관리, 데이터 수명 주기관리, 데이터 품질 관리에서 공유되도록 한다.

셋째 영역인 'Ⅲ. 관리 환경 구축'에서는 선행 영역에서 수립한 이행계획이 효율적으로 진행될 수 있도록 지원하고, 계획의 진행현황을 파악할 수 있는 정보를 수집하여 관리한다.

- (3.1 Definition of KPIs) DG 프로그램의 이행에는 인력, 절차 그리고 기술이 필요하다. 이중에서 인력과 절차는 무형이어서 평가하여 관리하지 않으면, DG 프로그램이 체제적이고 효과적으로 진행되기 어렵다. 평가지표(KPIs: Key Performance Indicators)는 이러한 이유로 필요하다. 그러므로 KPIs는 제안

DGMM의 구성요소를 기준으로 비즈니스와 기술 부분을 나누어 정의하고, 이를 측정할 수 있는 대시보드를 구축하여 활용하도록 한다.

- (3.2 Security & privacy rule management) 보호 대상 정보자원(서버, 스토리지 등)이나 개인정보를 대상으로 하는 법령이나 제도를 점검하여 체계화하여 기업 내부의 행동 규정 등을 정의하고 관리하는 절차이다.

마지막 영역인 'Ⅳ. DG 구현'에서는 실질적인 DG 프로그램을 이행하는 영역이며, 총 3개의 절차가 있다.

- (4.1 Master data management) 마스터 데이터는 자주 변하지 않고, 기업 내의 여러 정보시스템에서 공유되는 기준 데이터(또는 핵심 데이터)의 집합을 말한다. 그러나 이러한 중요성에도 불구하고 마스터 데이터는 전사적으로 비즈니스 프로세스, 시스템, 응용 프로그램 등에서 복제 또는 분산되는 경우가 종종 발생한다. 이로 인하여 데이터의 중복이나 불일치와 같은 문제가 발생한다. 이를 해결하기 위하여 마스터 데이터 관리(MDM)가 필요하다. MDM은 기업 등이 비즈니스 전반에 대하여 기준 데이터를 식별, 마스터 데이터로 정의하고 통합하여 관리하는 것을 말한다. 여기에서 요구되는 세부사항은 데이터의 프로파일링과 분류, 마스터 데이터 설계, 데이터 매핑, 데이터 통합 등이 있다. 이 마스터 데이터에 개인정보가 포함되어 있을 경우에는 업무 내용과 공동 이용의 근거를 명확히 확인하고 관리·이용하여야 한다. 그러하지 않으면, 개인정보의 목적 외 이용에 해당하여 위법할 수 있다.
- (4.2 Data life-cycle management) 데이터의 생성(또는 수집)에서부터 소멸까지를 관리하는 것은 데이터 수명주기관리(DLM)라 한다. 이는 기업 등의 모든 데이터를 비즈니스 관점의 가치기준으로 분류하고, 정책·프로세스·도구를 동원하여 비용대비 효과를 최적화하는 절차이다. DLM을 위한 대표적인 정책으로 비즈니스 가치가 줄어든 비활성 데이터를 마스터 데이터에서 분리하여 별도로 관리(또는 삭제)함으로써 저장장치의 비용절감과 시스템 성능을 향상시키는 것이다. 이를 위한 세부 단계는 데이터 흐름관리, 분리(또는 파기)대상의 식별, 파기 등이 있다. 또한 개인정보는 당초의 수집 목적을 달성한 경우에는 즉시 파기해야 하므로 이를 위한 절차도 포함하여 함께 관리하여야 한다.
- (4.3 Data quality management) 기업 등이 부문별, 업무별로 별도의 정보시스템을 구축·운영할 경우에 데이터의 중복과 불일치 등의 문제가 발생하게 된다. 이러한 현상을 완화하거나 제거하기 위한 활동을 데이터 품질관리라 한다. 이를 통해 데이터의 정확성·일관성·적시성 등을 확보하는 절차이다. 세부적인 내용으로는 품질진단, 품질 측정, 품질 모니터링, 데이터 정화(크린징), 데이터 계보 및 임팩트 분석 등이 있다. 개인정보의 경우는 앞에서 논의한 바 있는 8원칙의 두 번째인 "정확성의 원칙"에 해당하므로 개인정보의 품질관리도 당연히 포함되어야 한다.

4.6 제안 DGMM과 종합이행절차를 개인정보 활용과 보호에 적용한 예

기업 등은 데이터의 관리를 위해 DG 프로그램을 도입하고자 한다면, 제안 DGMM의 사용을 우선 고려할 수 있다. 이 경우에 Fig.8의 종합이행절차에 따라 모든 절차를 진행하여야 하지만, 본 연구에서는 여러 절차 중에서 '1.2 Assessment of DG maturity level(DG 성숙도 평가), 제안 DGMM의 (8) Use of Personal Data(개인정보 활용)' 요소를 사례로 적용해 본다.

기업 등은 비즈니스와 관련하여 개인정보의 수집·이용이 필요하면, 법 제33조의 개인정보 영향평가, 제15조와 제18조의 개인정보 수집과 제3자 제공, 그리고 법 제29조 안전조치의무 등 관련 규정에 따라 개인정보의 활용과 보호에 대한 기업 등의 내부 정책방향과 세부이행내용, 그리고 인력과 조직 등이 포함된 '개인정보의 내부관리계획'을 법 제29조 및 「개인정보의 안전성 확보조치 기준」 제4조에 따라 추진하여야 한다. 이때 만약 이때 '보호'에만 중점을 둔다면, 어떤 데이터가 개인정보에 해당하는지를 명확하게 알 수 없는 상황에서 개인정보 암호화 등 정보보호 관련 기술을 적용함으로써 관련 법령에서 정하고 있는 개인정보 수명주기 전반에 걸친 법령상의 규정을 준수하지 못할 수 있다. 이 경우에 개인정보보호 담당자는 사후에 이를 수습해야 하지만, 개인정보를 이용하는 업무 관계자를 이해시키고 설득하는 것은 쉬운 일이 아니다. 이러한 문제를 사전에 DG 차원에서 해결하는 것이 바람직한 방안이 될 수 있다.

이러한 상황에서 기업 등이 개인정보의 활용과 보호를 고려하여 DG 프로그램의 도입을 '제안 DGMM'과 '종합이행절차'에 따라 진행할 경우에 Fig.8의 '1.2 DG 성숙도 평가', 제안 DGMM의 '(8) 개인정보 활

용' 요소에 대하여 Table 4의 내용을 참고하여 관련 정보를 수집·분석하여 성숙도 수준을 평가한다. Table 4는 보호법, 동법 시행령 및 고시에서 정한 의무사항을 예로 정리한 것이며, 현업에서 적용하려면 전문기관 등의 도움을 받아 보다 구체적으로 정의하여 활용하는 것이 좋다. 만약 기업 등이 영리 목적으로 정보통신망을 이용할 경우는 정보통신망법, 그리고 동법 시행령 및 고시에 따른 별도의 기준을 Table 4를 참고하여 새로이 정의하여야 한다. 여기서 Table 4를 활용과 보호로 구분하여 보면, '1.수집'과 '2.이용·제공'은 활용 관점으로, '3.저장·관리'는 보호 관점으로, '4.파기'는 공통 관점으로 대략적인 구분이 가능하다.

DG의 성숙도 평가를 위해 수집하는 정보는 앞에서 논의한 바 있는 설문정보, 해당 기업 등에 요청하여 제공받은 정보, 그리고 진단도구로 시스템에서 자동으로 수집한 정보 등이 있다. 이를 종합·분석하여 DG 관점에서 현재의 '(8) 개인정보 활용'의 수준을 평가한다.

이때 중요한 것은 대상 기업 등이 수집·이용하고 있는 수많은 데이터 속에서 개인정보를 식별하여 관리하고 있는지, 그리고 부족한 것이 무엇인지 알아보는 것이다. 또한 수집·이용 중(또는 예정)인 데이터 속에서 개인정보를 어떻게 식별하고, 그리고 그 식별에 필요한 메타데이터를 어떻게 수집하고 관리하는가이다.

Fig.9는 'KKKK'라는 이름의 DB Table, 즉 DF에이며 (E.3)와 같이 표현할 수 있다.

KKKK = {AAA,BBB,CCC, … , HHH} -------- (E.3)

여기서 AAA, BBB 등과 같은 항목의 특성을 알 수 없기 때문에 KKKK가 PDF에 해당하는지 여부를 판단할 수 없다. 각 항목의 특성을 알 수 있는 방법은 정보를 수집할 때, 항목의 특성을 미리 정하는 방법과 사후에 수집한 데이터의 실제 값으로부터 특성 정보를 찾아내는 방법 등을 생각할 수 있다. 이러한 방법으로 Fig.9의 'Meta-Data'의 확보가 가능하다. 예에서 'AAA' 항목은 'Name(이름)'이고, 한글로 3-4자로 구성된다는 것을 알 수 있다. 과거에는 'AAA'를 'Name'과 같이 항목이름을 사용함으로써 관계자가 항목 속성을 항목이름으로 추정이 가능하였지만, IoT 등 데이터 수집 도구가 다양해짐으로써 그 방법으로는 한계가 있어 별도로 수집하여 메타데이터로 관리할 필요가 있다. 그리고 이러한 정보를 이용하여 Fig.5에 있는 DI나 QI를 알아내고, Table 2를 참고하여 KKKK가 PDF의 해당여부를 판단할 수 있다. 물론 예에서 KKKK는 BBB가 DI(주민등록번호)이므로 PDF에 해당한다.

Table 4. References for DG maturity assessment for use and protection of personal data

1. Collection(수집) • It first generates (E.1) or (E.2), and whether the items in the result correspond to a direct identifier(DI) or a quasi-identifier(QI). • If an item of an arbitrary DB contains DI or two or more QIs, the DB is a PDF, and the value stored therein is defined as personal data. • Check the legal basis of collection and use, such as consent forms, for the item values in the PDF.
2. Storage & Management(저장 · 관리) • Check whether the resident registration number is encrypted. • Identify access rights related materials such as HR documents. • Check the appropriateness of technology for access control and monitoring of access to personal data. • Confirm whether or not to prevent the personal data access record from being stored, altered, or damaged. • And other procedures for guaranteeing rights for owners of personal data.
3. Use & provide(이용 · 제공) • Confirm consent or legal justification for the purpose of use or the provision of third parties. • Identify de-identification processing methods and process for providing personal data to third parties.
4. Destruction(파기) • Check the criteria for confirming the achievement of the purpose of personal information at the time of collection. • Check how to identify personal data to be destroyed or deleted. • Check personal data and methods to be kept in accordance with laws and regulations.

이상과 같은 작업의 예는 실제 DG 프로그램을 도입하는 과정에서는 Fig.8의 2.4에 해당한다. 그러나 '1.2 성숙도 평가'에서는 Fig.7 DGMM의 '(8) 개인정보 활용'에 해당하며, 평가대상 기업 등에서 위의 작업이 어느 수준까지 완성되어 있어 있는가를 수집한 관련 정보로 평가하게 된다.

성숙도 평가 결과는 Fig.10과 같은 형태로 표현할 수 있다. Fig.10에는 평가대상 기업 등의 현재 수준은 'Initial(1단계)'이고, 미래의 최종 기대 목표 수준은 'Defined(3단계)'로 평가되었다. 이러한 평가결과는

(Table Name) KKKK

AAA	BBB	CCC	DDD	EEE	FFF	GGG	HHH
G.Park	590600	1123-	D.Gu	500	Apt.	10G	30Min.
G.Hong	601010	2324-	K.Gu	450	Villa	8G	20Min.
G.Jung	751011	3333-	G.Gu	370	-	30G	20Min.
G.Cho	801020	4535-	M.Gu	250	-	30G	50Min.

(Meta-Data) : Data for identifying personal data identifier

```
AAA = "name"                    / Name of the person, 3-4 characters in Korean/
BBB = "resident registration No." / Korean identification information, 13 digits/
CCC = "Tel No."                 / Telephone number, 11 digits/
DDD = "address"                 / sido, si-gun-gu, dong/
EEE = "salary"                  / number : NN.NNN.NNN.NNN/
FFF = " real estate"            / Apartment, Villa, House, etc. /
GGG = "data usage"              / number : NNN(GB)
HHH = "Phone call time"         / time : NNN.NN
```

Fig. 9. Examples of DF and Metadata

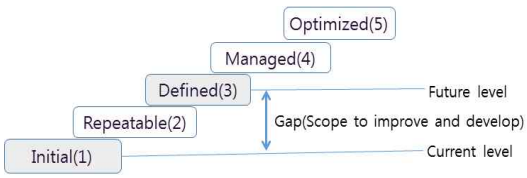


Fig. 10. A result of evaluating the maturity level of the 'Use of Personal Data'

Fig.8의 1.4에서 최종 목표 달성을 위한 과제를 발굴하고, 과제별 로드-맵의 작성에 활용한다.

V. 결론 및 향후과제

5.1 결론

지금까지 DG와 DG 프레임워크에 대하여 살펴보고, 전사적으로 분산·저장되어 있는 데이터를 대상으로 PDF와 개인정보를 식별하는 방법을 제시하였다. 그리고 개인정보의 활용과 보호를 위한 새로운 DG 프레임워크인 '제안 DGMM'과 DG 프로그램을 위한 '종합이행절차'도 제안하였다.

일반적으로 일정 규모 이상의 기업 등은 대부분 개인정보를 수집·이용하고 있다. 그 이유는 인사 데이터, 고객 데이터 등을 필수적으로 수집·이용하기 때문이다. 또 다른 이유는 센서, IOT 등 4차 산업 관련 다양한 기술로 수집한 데이터는 직접 식별자(IP, MAC 등)를 포함하고 있어 DF나 DF1, 그리고 Table 2에 따라 PDF에 해당하기 때문이다.

이와 같이 기업 등이 수집·이용하는 개인정보를 포함한 대량의 데이터는 전사적으로 공유·활용되어야 비즈니스 이익 극대화에 도움이 된다. 그러기 위해서는 데이터 품질 등의 확보를 위하여 DG 프로그램의 도입이 고려되어야 한다.

본 연구에서 제안한 내용은 최근 국내·외에서 강화되고 있는 개인정보 보호 여건 속에서 기업 등이 자사의 DG 성숙도의 수준을 평가하고, 그 결과를 기초로 DG 프로그램 도입 여부의 판단과 추진 방향의 설정 등에 많은 도움이 될 것으로 기대한다.

5.2 향후 과제

제시한 PDF 식별방법, '제안 DGMM'과 '종합이행절차'는 DG 프로그램 도입의 방향설정에 필요한 기본적인 것이다. 하지만, 이를 보다 체계화하고 구체화하기 위하여 추가로 필요한 연구는 아래와 같이 네 가지를 들 수 있다.

첫째는 (E.1)의 각 항목속성을 직접 식별자, 간접 식별자, 민감 정보, 일반정보로 분류하는 방법이 법령 등에서 명확히 정하고 있지 않아 인위적으로 판단하는 방법 밖에 없다. 만약 연구를 통하여 이를 자동으로 분류하여 정확도를 높일 수 있다면, 개인정보의 활용과 보호 측면에서 많은 도움이 될 수 있다. 그 이유는 인위적으로 항목속성의 분류결과를 자동분류로 통해 검증할 수 있다면 보다 정확성을 높일 수 있기 때문이다. 또 다른 이유는 데이터베이스 설계 초기에 해당 항목에 어떠한 내용이 저장될지를 알 수 있어 인위적인 판단이 가능하지만, 사후에 관련 문서의 손상이나 관계자 퇴사 등으로 그 내용을 정확이 알 수 없을 경우, 즉 데이터베이스 내에 저장되어 있는 실제 값 외에 관련 다른 정보가 없는 경우에 이를 이용할 수 있기 때문이다.

둘째는 효율적인 DG 평가지표(KPIs)의 개발과 모니터링 방안의 연구이다. DG 프로그램이 성공하기 위해서는 인력, 절차, 그리고 기술이 체계적, 유기적으로 동작하여야 한다. 이의 동작여부를 판단할 수 있는 평가지표의 개발과 이를 쉽게 알 수 있는 모니터링 할 수 있는 도구를 연구하여 개발되어야 한다.

셋째는 유럽연합의 GDPR에만 있는 데이터 이동권(data portability)[13] 등을 제안 DGMM과 절차에 확대하는 연구이다. 2018년 5월부터 시행중인 GDPR이 국내에 미치는 영향이 어느 정도 파악이 되면, DG 프로그램 관련 기업은 적절한 대처가 필요하다. 이를 대비하여 제안 DGMM과 종합이행절차의 변경이나 개선이 필요할 수도 있기 때문이다.

넷째는 제안 DGMM과 종합이행절차를 데이터 거

버넌스 현장에 적용하는 과정에서 문제점이 발견되면, 이를 개선하여 진화·발전시키는 것이다. 본 연구는 메타 데이터 관리, 데이터 품질관리, 데이터 통합, 마스터 데이터 관리 등 각각의 프로젝트 경험을 기초로 이루어졌고, 또한 연구결과의 실험을 위한 적절한 환경을 마련하기 어려워 실질적인 실험을 할 수 없었다. 따라서 본 연구의 결과를 현장에 적용하는 과정에 발견되는 문제는 수시로 적용하여 개선되어야 한다.

References

[1] Z. Panian, "Some Practical Experiences in Data Governance", World Academy of Science, Engineering Technology 62, pp. 939-946, 2010.

[2] G. Thomas, "The DGI Data Governance Framework", Data Governance Institute, Orlando, FL (USA), 2006.

[3] Sunil Soares, *The IBM Data Governance Unified Process*, MC Press, Sept. 2010.

[4] Seok-Soo Kim, "A Case Study of Implementation Data Governance for Enterprise Architecture", *Journal of Information Technology and Architecture*, 8(3), pp. 255-265, Sept. 2011.

[5] J. Ladley, *DATA GOVERNANCE : How to Design, Deploy, and Sustain an Effective Data Governance Program*, Morgan Kaufmann Press, 225 Wyman Street, Waltham, MA 02451, 2012.

[6] Kim, Sunho and Lee, Changsoo, "A Master Data Quality Management Framework, *Entrue Journal of Information Technology*, 9(2), pp. 109-121, July 2010.

[7] Choi, Wan Il, "Establishing Data Governance in the Publication and Use of Public Sector Information.", National Information Society Agency, pp. 319-323, 2011.

[8] K.Wende, "A Model for Data Governance -Organising Accountabilities for the Quality Management", *18th Australian Conference on Information Systems,* pp. 417-425, Dec. 2007.

[9] Kyoung_Ae Jang and Woo-Je Kim, "A Level Evaluation Model for Data Governance", Journal of *The Korean Operations Research and Management Science Society,* 42(1). pp. 66-77, 2017.

[10] Mark C. Paulk, Bill Curtis, M.B. Chrissis, and C.V. Weber, "Capability Maturity Model, Version 1.1", *IEEE Software*, Vol.10, no. 2 pp. 18-27 July 1993.

[11] Mark C. Paulk, Bill Curtis, and M.B. Chrisis, "CMMI Guidlines for Process Integration and Product Improvement, Addison-Wesley Long Publishing Co., 2003

[12] Y.J. Won, W.P.Park, J.Kwak, H.J.Kim, *The 4th Industrial Revolution & Information Security*, Chung Ram Press, Mapo-Seoul, 2018.

[13] P. Voigt and A. von dem Busche, *The EU General Data Protection Regulation, : A Practical Guide*, Springer Inter - national Publishing, 2017.

[14] Office for Government Policy Coordination, Ministry of the Interior and Safety ect. *Guidelines for De-identification of personal data,* 2016.

[15] K.E. Emam, B. Malin, "Appendix B : Concepts and Methods for De-identi -fying Clinical Trial Data.", National Academy of Sciences, 2015.

[16] Hwang Sooha and Kim Jeungduk, "A study on the goals and processes of privacy governance", *Review of KIISC,* 23(6), pp. 7-11, Aug. 2011,

[17] Ministry of the Interior & Safety, KISA, *Policy commentary for securing Personal Data safety*, 2015.

〈저자소개〉

이 영 상 (Lee YoungSang) 정회원
1986년 2월: 경북대학교 전자공학(전산) 졸업
1989년 2월: 미시건 주립대학 전자공학 석사
2003년 3월: 한국과학기술원(KAIST) 전자공학 박사과정
1989년 9월~2001년 8월: 대우전자, 대우통신, ㈜드림정보, 코스트코리아(이사) 등
2001년 9월~현재: ㈜데이터스트림즈 대표이사
2010년 1월~2012년 2월: 한국 S/W전문기업협회 회장
2013년~현재: 한국빅데이터학회 부회장
〈관심분야〉 데이터 품질관리, 데이터 거버넌스, 빅데이터

박 원 환 (Park WonHwan) 정회원
1982년 2월: 충남대학교 계산통계학과 졸업
1986년 2월: 충남대학교 계산계학과 석사
2002년 2월: 순천향대학교 전산학 박사
1988년 2월~2005년 2월: 통계청
2005년 3월~2011년 9월: 정보통신부, 행정안전부 정부통합전산센터
2011년10월~2016년10월: 개인정보보호위원회 조사과장, 조사조정관
2016년12월~현재: 충남대학교 핀테크보안연구센터
〈관심분야〉 개인정보보호, 데이터 거버넌스, 정보보호, 빅데이터

신 동 선 (Shine DongSun) 정회원
1985년 2월: 충북대학교 수학과 졸업
1995년 7월: 한국생산성본부 전문위원
2009년 2월~현재: ㈜데이터스트림즈 거버넌스 총괄 상무
〈관심분야〉 데이터 거버넌스, 데이터 품질관리

원 유 재(Won YooJae) 종신회원
1985년 2월: 충남대학교 계산통계학과 졸업
1987년 2월: 충남대학교 계산계학과 석사
1998년 8월: 충남대학교 전산학과 박사
1987년 2월~2001년 2월: 한국전자통신연구원 책임연구원(팀장)
2001년 3월~2004년 8월: 안랩유비웨어, 안랩 CTO
2004년 9월~2014년 2월: 한국인터넷진흥원 본부장
2014년 2월~현재: 충남대학교 컴퓨터융합학부 교수
〈관심분야〉 정보보호, 사이버보안, 네트워크 및 이동통신 보안