

개인정보처리방침의 데이터를 활용한 개인정보보호 현황 분석

이재근,^{1*} 강상욱,^{1‡} 엄흥열²
¹한국정보화진흥원, ²순천향대학교

Analysis of Personal Information Protection Circumstances based on Collecting and Storing Data in Privacy Policies.

Jae-geun Lee,^{1*} Sang-ug Kang,^{1‡} Heung-Youl Youm²
¹National Information Society Agency, ²SoonChunHyang Univ.

요약

개인정보보호 분야는 타 분야에 비해 통계 및 현황에 대한 정보가 부족하다. 또한 개인정보보호 분야는 정보화, 정보보안, 법 등 어느 분야에도 일방적으로 속하지 않아 부수적으로만 실태 파악이 되고 있어 앞으로도 한계가 있을 것이다. 본 논문에서는 약 11만 개의 개인정보처리자가 공개하고 있는 개인정보 처리방침을 이용하여 방대한 양의 데이터를 수집하고 저장하여 이를 분석할 수 있는 방안을 제시하고 실제 결과를 보여준다. 제시된 방안은 데이터의 품질이 보장되고 시간과 비용을 절감할 수 있어 기존 방식에 비해 결과에 대한 정확성과 현황 파악에 대한 신속성에 있어 유리하다. 또한 다양한 분석이 가능하여 개인정보보호에 대한 실태 파악의 새로운 돌파구 역할을 할 수 있다. 본 논문에서는 개인정보보호와 빅데이터 분야의 접목을 통해 개인정보보호의 발전 방안을 제시하고 있다.

ABSTRACT

A field of privacy protection lacks statistical information about the current status, compared to other fields. On top of that, since it has not been classified as a concrete separate field, the related survey is only conducted as a part of such concrete areas. Furthermore, this trend of being regarded as a part of fields such as informatization, information protection and law will continue in the near future. In this paper, a novel and practical way for collecting and storing a big amount of data from 110,000 privacy policies by data controller is proposed and the real analysis results is also shown. The proposed method can save time and cost compared with the traditional survey-based method while maintaining or even advancing the accuracy of results and speediness of process. The collected big personal data can be used to set up various kinds of statistical models and they will play an important role as a breakthrough of observing the present status of privacy information protection policy. The big data concept is incorporated into the privacy protection and we can observe the method and some results throughout the paper.

Keywords: Privacy Policy, Personal Information Protection, Data Analysis, Statistics

I. 서 론

우리나라에서는 2011년 9월 30일부터 개인정보보호법이 시행되었고 2011년 10월에 조사된 결과에 의하면 국민의 98.5%가 개인정보보호의 중요성을 인식하고 있다[1]. 미국에서 실시된 프라이버시 관련 설문 조사에서도 응답자의 82%가 개인정보를 수집하는 스마트폰 또는 인터넷 회사에 대해 걱정스럽다(concerned) 라고 응답했다[2]. 이처럼 개인정보에 대한 일반인의 인식 수준이 높아지고 개인정보 침해에 대한 우려도 높은 상황에서 기술적인 측면과 관리적인 측면에서 다양한 대책이 시도되고 있다. 기술적으로는 개인정보의 노출 및 유출방지에 주로 초점을 맞추고 암호화, 접근제어, 모니터링 등의 분야에서 다양한 기술이 선보이고 있으며 관리적인 면에서는 조직이 준수해야 하는 정책 및 내규, 내부감사 활동, 활동 결과의 준거 여부에 대한 상시 모니터링, 개선 활동 체계 등에 대한 많은 연구가 진행되고 있다[3].

우리나라에서 개인정보보호와 관련된 현황 및 실태의 파악은 설문조사방식에 의하여 실시되고 있다. 한국정보화진흥원에서 매년 실시되는 정보화 통계조사에서는 개인정보보호와 관련하여 2011년까지 '홈페이지 개인정보 수집 현황', '개인정보보호지침 제정·운영 현황', '개인정보 유·노출 사고 피해 경험', '피해신고 및 피해 복구비용 발생' 등 5개 항목을 조사하였으나[4], 2012년 조사에서는 '홈페이지 개인정보 수집 현황', '개인정보보호지침 제정 및 운영 현황'을 제외한 3개 항목만 조사하고 있고[5], 한국인터넷진흥원의 2012년 정보보호 실태조사에서도 개인정보보호 관련 항목이 45개에 이르고 있으나, 2007년부터 조사되었던 '개인정보 취급 방침별 공개 여부', '개인정보 수집, 이용/제공시 이용자 동의 확보 여부', '수집한 개인정보의 제3자 제공/취급 위탁 여부' 등 개인정보처리자의 개인정보 수집, 제공 및 관리와 관련된 7개 항목에 대한 조사가 이루어지지 않고 있다[1][6].

기업이나 국가에서 개인정보에 대한 관리를 제대로 수행하기 위해서는 정확한 현황 파악과 이를 바탕으로 다양한 분석이 전개되어야 하지만 이는 많은 노력과 비용이 필요하다. 이러한 문제를 해결하기 위해서 법적으로 모든 개인정보처리자가 공개해야 하는 개인정보처리방침(privacy policy)을 활용해 개인정보보호에 대한 의미 있는 통계를 도출하고 이를 분석 할 수 있다면 기존의 부족한 현황 자료를 보완하고 실제 적용도 간편하여 정책 개발 및 빅데이터 분석에 다양하

고 편리하게 활용할 수 있다. '개인정보보호법' 상에는 '개인정보처리방침'으로 '정보통신망 이용촉진 및 정보보호 등에 관한 법률(이하 정보통신망법)' 상으로는 '개인정보취급방침'으로 각기 달리 표기되지만 본 논문에서는 개인정보보호법을 기준으로 서술하고 이하에서는 처리방침으로 통칭하고자 한다.

처리방침을 현황 파악에 활용한 사례로는 소비자원이 개인정보 고위험군 사업자의 문제점을 분석하고 개선 방안을 마련하기 위해, 이동통신서비스업, 신용카드업, 생명 및 화재 보험업 등 6개 업종 60개 사업자의 처리방침 내용을 조사하고 해당 업종의 개인정보 수집, 제공, 관리 실태를 분석한 바 있다[7].

일부 언론사에서는 홈페이지에 공개된 처리방침을 이용하여 주민번호를 비롯한 개인정보가 제3의 많은 개인정보처리자에게 제공 또는 위탁되고 있다는 기사를 보도하기도 하였다[8][9].

처리방침에서 공개한 내용을 자동으로 수집하고 분석하여 개인이 설정한 개인정보보호 수준과 비교하여 개인의 요구에 부합되는지 여부를 판단하는데 활용하기도 한다. AT&T가 무료로 제공하는 Privacy Bird라는 웹 브라우저용 프로그램은 P3P (Platform for Privacy Preferences)[10]로 작성된 처리방침을 파싱하고 읽을 수 있는 기능을 탑재하여 한 인터넷 사이트에 접속할 때 P3P로 입력되어 있는 사이트의 처리방침을 자동 수집한다. 이 정보와 미리 입력한 개인의 프라이버시 성향을 비교 후에 사이트 접속자에게 적합, 부적합 여부를 알려준다. P3P는 처리방침을 XML을 이용하여 정형화된 형식으로 표현하여 소프트웨어 에이전트가 자동으로 검색하고 해석할 수 있도록 도와준다. 1997년부터 W3C 주도로 AOL, HP, 마이크로소프트 등 업계와 시민단체가 참여하여 진행하였으며, 2001년 개발이 완료된 후 시험 운용 끝에 2002년 4월에 국제표준으로 승인되었다[10][11].

마이크로소프트사의 웹브라우저인 인터넷 익스플로러에서도 P3P를 해석하는 기능을 탑재하여 사용자가 개인정보를 자율적으로 보호할 수 있도록 쿠키를 선별적으로 받아들이거나 거부할지를 사용자가 미리 지정한 쿠키에 대한 설정과 비교하여 결정한다. P3P의 설정은 인터넷 옵션의 컨트롤 패널에서 개인정보 탭을 통하여 설정할 수 있다. Cranor 등[12]은 P3P로 처리방침이 구현된 사이트를 대상으로 인기도 별로 건강 및 의료정보, 금융정보의 수집 및 마케팅 활용 여부 등의 정보를 분석하여 전체적인 프라이버시 보호 수준

을 제시한다. 예를 들어, 구글(google) 검색엔진에서는 10대 인기 검색 순위에 31.8%의 확률로 최소한 하나의 P3P를 채택한 사이트가 존재하고 프라이버시 보호 수준이 '낮음'으로 평가된다는 등의 결과를 보인다. 하지만 이러한 활용 사례는 낮은 P3P 채택률 때문에 한계를 보이고 있고 P3P가 표준으로 제정된 지 꽤 긴 시간이 지났지만 전체 웹 사이트의 10% 남짓만 P3P를 채택하고 있다[12][13].

이와 같이 기존의 개인정보보호 관련 현황에 대한 조사 방법은 설문조사나 P3P 해석에 기반하고 있기 때문에 일반적인 처리방침을 이용한 자료 수집 및 분석을 위한 정형화된 메타데이터 형식과 엔코딩 스킴을 제시하지 않았고, 국내외에서도 이에 대한 연구는 전무하다.

본 논문에서는 채택률이 저조한 P3P 사용 여부에 관계없이 1,000개 개인정보처리자의 처리방침 및 연관 데이터를 정형화된 메타데이터 형식 및 엔코딩 스킴을 통해 수집하고 이를 통계적으로 분석한다. 현재 부족한 개인정보보호 관련 통계와 현황을 보다 풍부하고 정확하게 적은 인력과 시간을 투입하여 알 수 있는 새로운 방법을 제안하여 프라이버시 관련 빅데이터 분석의 기반을 제시한다. 참고로, 개인정보보호법에 정의된 용어인 '정보주체'와 '개인정보처리자'를 그대로 차용하여 이 논문 전체에서 사용한다.

II. 데이터의 수집 및 저장

2.1 처리방침 데이터의 품질 요소

처리방침이란 개인정보처리자가 개인정보 처리기준 및 보호조치 등을 문서화하여 제3자에게 공개하는 공식체도로, [표 1]에서 나열하고 있는 ①번부터 ⑫번까지의 정보를 포함하고 있다. 개인정보보호법 시행령 제31조에 의하면 처리방침은 우선적으로 인터넷 홈페이지에 공개하도록 의무화 되어 있다. 또한 개인정보보호법 제30조에 따라 개인정보 수집 항목 및 목적을 변경하거나, 업무 제휴 등 기업 전략 변경에 따라 개인정보 제3자 제공 대상을 변경하는 경우, 개인정보 처리 위탁회사를 변경 또는 추가하는 경우, 개인정보 책임자를 변경하는 경우 등 개인정보 처리와 관련된 내용이 변경되는 경우에도 처리방침을 수정하여 공개하도록 하고 있다. 따라서 처리방침을 통해 알 수 있는 개인정보보호와 관련된 정보는 풍부하고 안정적이며 지속적으로 획득 할 수 있는 데이터 자원으로서 이

[표 1] 직접 및 연관 데이터 항목

구 분	수집 가능 데이터	비 고
처리방침 직접 데이터	①개인정보 수집 목적	필수기재 사항 (법제30조)
	②개인정보 처리 및 보유 기간	
	③제공 목적, 제공 사업자, 제공 사업자 수, 제공기간	
	④위탁 목적, 위탁 사업자, 위탁 사업자 수	
	⑤정보주체의 권리의무 및 행사 방법 공개 여부	
	⑥개인정보 수집 및 이용 세부 항목	필수기재 사항 (시행령 제31조)
	⑦개인정보 파기 방법 공개 여부	
	⑧개인정보보호를 위한 안전성 조치 공개 여부	
	⑨개인정보책임자 공개 여부	필수기재 사항 (지침제37조)
	⑩방침 변경 관리에 관한 사항	
	⑪열람청구권 행사 방법 및 권 익침해에 대한 구제 방법 공 개 여부	임의기재 사항 (지침제38조)
	⑫처리방침의 명칭	필수 사항 (지침제34조)
개인정보 처리자 연관 데이터	⑬업종, 상시 근로자수, 소재지, 매출액 등	기업정보 제공 회사로부터 수집
홈페이지 연관 데이터	⑭로그인 유형, 실명인증 유형, 미성년자 정보 수집 여부, 개인정보 수집 유형 등	홈페이지에서 수집

를 적절하게 활용하는 것은 자연스럽게 중요하다.

처리방침에 포함되어 있는 정보가 빅데이터 처리에 필요한 활용성 높은 양질의 데이터가 되기 위해서는 고유품질(intrinsic), 맥락품질(contextual), 접근 품질(accessibility), 표현품질(representational) 등 네 가지 측면의 품질이 충족되어야 한다[14]. 처리방침을 활용할 가치가 충분한 지 데이터의 품질 측면에서 분석하고자 한다.

첫째, 고유품질은 데이터가 실제 값을 나타내며 편향되지 않고 사람들에게 사실이라고 받아들여져야 한다. 처리방침의 공개항목과 내용은 법과 시행령에서 규정되어 있고 정보주체에게 개인정보 처리에 대한 내용을 공개하는 약관 수준의 성격을 지니고 있다. 따라서 그 내용이 실제 내용을 반영하지 않는다면 정부의 조사 결과에 따라 과태료를 처분 받거나, 화면 캡처 등을 통해 각종 분쟁에 불리한 증거로 쉽게 사용되거나, 불공정한 개인정보처리자로 인식되는 등 실익이 없기 때문에 상당한 수준의 정확성이 있다고 판단할

수 있다. 또한 처리방침과 연계되어 존재하는 인터넷 홈페이지를 통해 개인정보 수집 항목 등 많은 내용을 실제 확인 할 수 있다는 점도 사람들의 데이터에 대한 신뢰성을 높일 수 있다.

둘째, 맥락품질은 데이터가 유효한 시간 내의 것이고 충분한 폭과 깊이의 연관성이 있어야 한다. 본 논문에서의 데이터 처리 목적은 개인정보보호에 대한 통계를 수집하고 현황을 파악하는 것이다. 처리방침에 있는 데이터는 개별 개인정보처리자의 개인정보 처리 내용을 12개의 항목에 걸쳐 일목요연하고 폭 넓게 공개한다. 한국정보화진흥원에 따르면 홈페이지에서 개인정보를 수집하거나 활용하는 개인정보처리자는 2011년에 약 12만 6천개로 이중 87.7%인 약11만개 개인정보처리자가 처리방침을 공개하고 있다(4). 이

는 개인정보 현황 파악이란 목적에 부합한다. 또한 법에 의하면 처리방침을 수립하거나 변경할 경우에도 공개하도록 되어 있어 상당한 수준의 최신성이 있다.

셋째, 접근품질은 데이터가 이용 가능해야 하고 쉽고 빠르게 획득할 수 있어야 한다. 처리방침은 해당 홈페이지의 시작페이지에 색깔 및 두꺼운 글씨로 눈에 띄게 표시하도록 법에 정해져 있다. 따라서 쉽게 처리방침을 찾을 수 있고 인력이나 인터넷 로봇을 통해 처리방침 데이터 획득이 용이하다. P3P로 처리방침이 작성되었을 경우 접근성이 보다 높아진다.

그리고 대상 데이터는 법에 의해 공개하도록 되어 있어 개인정보 책임자 등 일부 항목을 제외하고는 이용하는 데 따른 제약이 없다.

넷째, 표현품질은 데이터가 간결하고 일관되게 표

(표 2) 수집 데이터에 대한 정형화된 표현

구분	데이터 요소명	데이터 요소 설명	빈도수	데이터타입 (인코딩스킴)
①	수집 목적	처리방침에 기술되어 있는 개인정보 수집, 이용 목적	1..n	목적 Type
	수집 방법	처리방침에 기술되어 있는 개인정보 수집 목적별, 개인정보를 수집하는 방법	1..13	수집방법 Type
②	활용 기간	처리방침에 기술되어 있는 개인정보 이용 및 보유 기간	1..n	기간 Type
③	제공	개인정보 제3자 제공관련 내용 ※ 상위 요소	1..1	기술 Type
	*제공사업자 수	처리방침에 기술되어 있는 개인정보 제3자 제공 개인정보처리자의 수	1..1	Num Type
	*제공 목적	처리방침에 기술되어 있는 개인정보 제3자 제공 목적	1..n	목적 Type
	*제공 사업자	처리방침에 기술되어 있는 제3자 제공 대상 개인정보처리자의 명칭	1..n	Text Type
	*제공 기간	처리방침에 기술되어 있는 제3자 제공 목적별 개인정보 제공 기간	1..n	기간 Type
④	위탁	개인정보 위탁업무 관련 내용 ※상위 요소	1..1	기술 Type
	**위탁사업자 수	처리방침에 기술되어 있는 개인정보 위탁사업자의 수	1..1	Num Type
	**위탁 목적	처리방침에 기술되어 있는 개인정보 위탁사업의 목적	1..n	목적 Type
	**위탁 사업자	처리방침에 기술되어 있는 개인정보 위탁목적별 위탁사업자의 명칭	1..n	Text Type
⑤	권리의무 여부	권리의무 및 행사 방법이 처리방침에 기술되어 있는지 여부	1..1	Boolean Type
⑥	수집 항목	개인정보처리자가 수집 이용하는 개인정보이 세부 항목명	1..n	항목 Type
⑦	파기조치 여부	처리방침에 기술되어 있는 파기 절차 및 방법 기술 여부	1..1	Boolean Type
⑧	안전성조치 여부	안전성 조치가 처리방침에 기술되어 있는지 여부	1..1	Boolean Type
⑨	책임자 여부	개인정보책임자를 처리방침에 기술하고 있는 유형	1..1	기술 Type
⑩	방침제정일	처리방침에 기술되어 있는 처리방침 제정 날짜(년.월.일) 정보	0..1	Date Type
	방침변경 관리	방침 제정 및 변경 일자를 처리방침에 기술하고 있는 유형	1..1	변경관리 Type
⑪	열람청구 여부	열람청구 행사 방법이 처리방침에 기술되어 있는지 여부	1..1	기술 Type
⑫	처리방침 명칭	홈페이지 첫 화면에 표시되어 있는 처리방침의 명칭 코드	1..1	명칭 Type
⑬	업종 분류	개인정보처리자가 속해 있는 한국표준산업분류표에 따른 업종 코드	1..1	업종 Type
	상시근로자수	개인정보처리자가 상시적으로 고용하고 있는 직원의 수	1..1	Num Type
⑭	로그인 유형	홈페이지에서 정보주체가 로그인할 수 있는 방법 코드	1..4	로그인 Type
	실명인증 유형	홈페이지에서 정보주체가 회원 가입시 실명인증 할 수 있는 방법	1..7	인증 Type
	수집 유형	홈페이지를 통해 개인정보를 수집하는 기능 또는 메뉴 구성 방법	1..1	수집 Type
	미성년 수집 여부	홈페이지를 통해 14세 미만 정보주체의 개인정보를 구분하여 수집 여부	1..1	Boolean Type

* 제공의 하위 요소, ** 위탁의 하위 요소

현되고 무슨 뜻인지 쉽게 알 수 있어야 한다. 처리방침의 내용은 12개 항목별로 구분되어 있고 각 항목의 내용도 이해하기 쉽게 되어 있다. 처리방침 공개 형식 및 포맷 등에 대해서는 별도의 지침이 없고 처리방침에서 발췌할 대상이 되는 데이터 자체도 비정형적이다. 이는 사람이 필요 데이터를 수집할 때 주관적으로 작성하거나 자동적으로 수집할 때 정확성 및 효율성을 떨어뜨릴 수 있어 품질 저해 요인으로 작용한다.

표현 품질을 보완하기 위해 본 논문에서는 처리방침에 대해 메타데이터를 정의하고 메타데이터 항목 별로 인코딩 스키를 개발하여 데이터가 간결하고 일관되게 표현될 수 있도록 하였다.

2.2 수집 데이터 항목

처리방침 데이터를 조사하는 과정에서 수집할 수 있는 데이터는 직접 데이터와 연관 데이터로 구분된다. 직접 데이터는 해당 개인정보처리자가 개인정보를 어떤 목적으로, 어떻게 수집, 이용, 제공, 위탁하고 있는지 등의 내용을 처리방침에 기술한 것으로부터 직접 파악할 수 있는 데이터를 의미한다.

연관데이터는 다시 처리방침을 공개하고 있는 해당 개인정보처리자의 명칭, 업종, 종사자 수 등 기업정보 제공회사를 통해 획득 가능한 개인정보처리자 연관데이터와 로그인 유형, 실명인증 유형, 미성년자 정보 수집 여부, 개인정보 수집 유형 등 해당 처리방침이 공개되어 있는 홈페이지를 통해 파악할 수 있는 홈페이지 연관 데이터로 구분되며 세부 내용은 [표 1]에 기술되어 있다.

2.3 데이터의 정형화 및 허용값 설계

수집하고자 하는 데이터는 모두 일반적인 텍스트 형태 및 사람이 인지 활동을 통해 추출하는 비정형 데이터이다. 본 연구에서는 효율적인 데이터 처리를 위해 관계형 데이터베이스를 이용하기 때문에 데이터를 정형화해야 한다. 데이터의 수집, 저장 및 관리를 용이하게 하고 입력 오류의 최소화 및 통계 분석의 용이성을 도모하기 위해, 수집 가능한 데이터를 범주화 및 코드화하여 통계처리 및 정보 추출에 용이하도록 할 필요가 있다. 따라서 메타데이터 저장소 표준규격인 ISO/ IEC 11179(15)의 가이드라인에 따라 [표 2]와 같이 각각의 데이터를 정형화하여 설계하고 인코딩 스키에 따른 허용값을 [표 3]에서 제시한다.

[표 2]는 구분, 데이터 요소명, 설명, 빈도수와 인코딩스킴으로 구성되어 있다. 예를 들어, '구분'은 [표 1]과의 연계를 위한 항목으로 ①은 법 제30조에 따른

[표 3] 인코딩 스키에 따른 허용값

데이터타입 (인코딩스킴)	허용값	설명
Text Type	텍스트(Text)	문자열로 구성된 텍스트
Num Type	숫자(Numeric)	정수
Boolean Type	Yes(1), No(0)	해당사항 여부
Date Type	년-월-일	개인정보처리방침 제정 일자
업종 Type	A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S	한국표준산업분류의 업종별 분류 코드
명칭 Type	개인정보처리방침, 개인정보취급방침, 기타	처리방침 표현방법 코드
목적 Type	[표 4] 목적Type 코드표 참조	수집, 제공, 위탁 목적 분류 코드
수집방법 Type	미기술, 홈페이지, 서면양식, 팩스, 전화, 상담게시판, 이메일, 모바일, 생성정보수집, 협력회사로부터, 고객센터, 시스템연계, 로그분석	개인정보 수집방법 코드
기간 Type	미기술, 1년, 2년, 3년, 5년, 영구, 준영구, 회원탈퇴시	이용기간 코드
항목 Type	이름, 영문이름, 생년월일, 주민등록번호, 여권번호, 외국인등록번호, 성별, 결혼여부, 기념일, 신체정보(혈액형 등), 장애급수, 별명, 닉네임, 국적, 사진, 우편번호, 주소, 전화번호, 휴대전화번호, 이메일, 이동통신사정보, 팩스번호, 아이디, 비밀번호, 비밀번호 질문/답변, 홈페이지, 아이핀(실명인증값), 가입인증정보, 위치정보, 직업, 현 학교/학과/학년, 현 직장명/직위, 학력사항, 경력사항, 병역사항, 자격사항, 성적, 카드번호, 계좌번호, 신용능력정보(재산, 납세실적 등), 신용도 판단정보(연체, 부도 등), 건강보험 납부액, 결제정보(결제승인정보 등), 서비스 이용 기록, 접속로그, 접속일시, 방문기록, 방문일시, PC사양정보, IP 주소, MAC 주소, 쿠키 정보, 마케팅수신동의, 이메일수신동의, SMS 수신동의, 전화수신동의, 차량번호, 운전면허번호, 관심사항, 취미, 법정대리인정보, 가족사항, 사업자등록번호, 상품권번호, 추천인 ID, 자기소개	수집하는 항목에 해당하는 명칭
기술 Type	미기술, 원격기술, 세부 내용 (제공사업자, 위탁사업자, 책임자) 기술	방침 내용 기술 형식
변경관리 Type	미기술, 현 방침의 제정일자만 공개, 방침 변경에 따른 버전관리	방침의 변경내용 기술 형식
로그인 Type	로그인 없음, ID/PW, 공인인증서, 아이핀	로그인 유형
인증 Type	실명인증 없음, 주민번호, 공인인증서, 아이핀, 핸드폰, 신용카드, 전자메일	실명인증 유형
수집 Type	회원가입, 게시판, 민원, 기타	개인정보 수집유형

처리방침의 필수 기재 사항이며 ‘수집 목적’이란 데이터 요소명으로 표현된다. 이 항목은 각 개인정보처리자별로 1개 이상의 수집 및 이용 목적이 존재하기 때문에 그 빈도수는 1..n 으로 표현할 수 있고 인코딩 스킴은 ‘목적 Type’을 써서 [표 3] 인코딩 스킴에 따른 허용 값에 따라 [표 4]의 값으로 제한한다.

[표 4]의 목적 Type 코드표는 방송통신위원회에서 2008년 5월 고시한 ‘개인정보취급방침의 전자적 표시방법’[16]의 수집 및 위탁 관련 목적 기호표에 기반하여 중복 또는 유사 목적의 코드는 통합하고, 제공 관련 목적을 보완하여 개인정보 수집, 위탁, 제공 목적에 공통적으로 사용할 수 있도록 재구성 한다.

데이터를 입력하고 관리하기 위한 프로그램은 마이크로소프트사의 액세스(Access) 소프트웨어를 이용하여 [그림 1]과 같이 GUI형태로 개발하였다. 데이터베이스 테이블 구현 시에는 통계분석을 용이하게 하기 위해 테이블 간에 중복 컬럼을 허용하고, 10개의 테이블로 분할된 데이터베이스 스키마들 간에는 기본 키를 포함하는 상위개체와 외래키를 포함하는 하위개체들과의 논리적 관계를 설계하고 구현하였다.



[그림 1] 데이터 수집을 위한 Access 프로그램

2.4 실 데이터의 수집

제안한 방법을 통해 다양한 빅데이터 분석을 할 수 있다. 그 중 처리방침의 제정 및 공시 현황과 처리방침 항목별 기술 현황의 정확성을 확인하기 위해 2011년 정보화통계조사[4]와 정보보호실태조사[1]의 기존 결과와 비교한다. [표 5]와 같이 정보화통계조사는 14,031개, 정보보호 실태조사는 5,042개의 표본을 대인면접조사 하였으나, 본 연구에서는 인력과 기간의 제한으로 1,000개의 표본을 제안된 방법의 가능성을 확인하는 차원에서 조사한다. 5명 미만의 고용인을 두고 있는 소상공인은 처리방침을 정확히 작성하고 공개

하기 어려울 것이라고 예상하고, 5명 이상을 고용하고, 네트워크가 구축된 사업체를 모집단으로 설정하는 정보보호 실태조사와 같이, 본 연구의 모집단은 홈페이지를 보유하면서 개인정보를 수집하는 종사자수 5명 이상의 개인정보처리자 69,837개로 설정하였는데, 이는 2011년 정보화통계집 354페이지의 ‘홈페이지 개인정보 수집 현황’의 수집 사례수 126,069개에서

[표 4] 목적 Type 코드표

코드	목적	적용업무
admin	웹사이트 및 시스템 관리	위탁
agency	기타 서비스 이행을 위한 대리점 운영	위탁
cert	회원제 서비스 이용에 따른 본인확인	수집, 위탁
complaint	불만처리 등 민원처리	수집, 위탁
content	개별 회원에 대한 맞춤 서비스 제공, 콘텐츠 제공	수집
delivery	물품배송 또는 청구지 등 발송	수집, 위탁
develop	신규서비스(또는 제품)개발 및 특화	수집, 위탁, 제공
finngmt	금융거래 본인 인증 및 금융 서비스	수집, 위탁, 제공
individual-analysis	접속 빈도 파악 또는 회원의 서비스 이용에 대한 통계 산출	수집
login	개인식별, 연령확인, 가입의사 확인, 불량회원 부정이용 및 비인가 사용 방지	수집
marketing	이벤트 등 광고성 정보 전달	수집, 위탁, 제공
other-purpose	기타 목적	수집, 위탁, 제공
payback	요금추심	수집, 위탁, 제공
payment	구매 및 요금 결제	수집, 위탁, 제공
proxy-agreement	만14세 미만 아동 개인정보 수집시 법정 대리인 동의여부 확인	수집
statement	고지사항 전달	수집, 위탁
warranty	AS 서비스 제공	위탁

[표 5] 모집단 및 조사 방법 비교

구분	제안방안	정보화통계 조사[4]	정보보호실태 조사[1]
모집단	홈페이지를 보유하며 개인정보 수집 하는 종사자수 5명 이상의, 개인정보처리자	전국 모든 사업체	종사자수 5명 이상의, 네트워크가 구축된 사업체
모집단수	69,837개	3,293,558개	420,134개
수집방법	인터넷 홈페이지 직접 조사	대인 면접 조사 (전화, 이메일 조사 병행)	
조사 표본수	1,000개	14,031개	5,042개
조사업종	13개 업종	13개 업종	13개 업종
조사 규모 구분	5~9명, 10~49명, 50~249명, 250명이상	1~4명, 5~9명, 10~49명, 50~249명, 250~999명, 1000명 이상	5~9명, 10~49명, 50~249명, 250명이상
모수 추정 오차	95% 신뢰수준 하에서 ±2.79%	95% 신뢰수준 하에서 ±1.3%	95% 신뢰수준 하에서 ±1.4%
조사시점	2013.4. ~ 5.	2011.4. ~ 6.	2011.10. ~ 12.

업종별 고용인수 1~4명에 해당하는 사례수를 제외한 5~9명, 10~49명, 50~249명, 250명 이상 구간의 사례수를 더하여 [표 6]의 업종별 '① 모집단'을 산출하고, 이를 바탕으로 업종별 '②비율'을 구한다[4].

그리고 3절에서 설명되는 기존 통계조사와의 결과 비교를 위해 1개 업종에 대해 집중해서 조사하기 보다는 기존 조사에서 조사한 13개 업종별 개인정보처리자 수와 조사대상의 고용 규모를 고려하여 전체업종과 고용규모별로 1,000개의 표본을 추출했으며 표본추출 결과는 [표 6]의 '③표본수'와 같다.

표본추출은 대한상공회의소 홈페이지[17]에서 제공하는 기업정보로 부터 업종, 종사자 수, 홈페이지 보유 여부를 살펴보고 무작위로 개인정보처리자 연관 데이터를 수집하였다. 그리고 조사대상의 처리방침 직접 데이터와 홈페이지 연관 데이터는 조사원 5명이 2013년 4월에서 5월초까지 6주일간 [그림 1]의 수집 프로그램을 이용하여 수집하였다. 수집된 데이터는 분석을 위해 엑셀파일로 통합 및 변환하고, 데이터의 정확성을 높이기 위해 입력 오류 등을 확인 및 보완하는 과정을 거쳤다.

[표 6] 처리방침 수집 및 분석을 위한 업종별 표본수

구 분	①모집단[4]		③표본수	④가중치
	사례수	②비율		
전 체	69,837	100.0%	1,000	69.8
1. 농림수산업	165	0.2%	2	82.5
2. 제조업	7,891	11.3%	118	66.9
3. 건설업	1,958	2.8%	31	63.2
4. 도매 및 소매업	7,232	10.4%	111	65.2
5. 운수업	729	1.0%	20	36.5
6. 숙박·음식점업	1,723	2.5%	35	49.2
7. 출판·영상·정보서비스업	3,439	4.9%	62	55.5
8. 금융·보험업	1,623	2.3%	36	45.1
9. 부동산 임대업	2,314	3.3%	34	68.1
10. 전문과학·기술서비스업	2,152	3.1%	31	69.4
11. 시설·사업지원 서비스업	1,355	1.9%	19	71.3
12. 협회·단체, 개인서비스업	4,995	7.2%	87	57.4
13. 기타	34,261	49.1%	414	82.8
5~9명	24,798	35.5%	305	81.3
10~49명	31,844	45.6%	324	98.3
50~249명	11,593	16.6%	189	61.3
250명 이상	1,602	2.3%	182	8.8

2.5 모수 추정

본 논문에서 제안한 처리방침 직접 조사방법의 타당성을 확인하기 위해 2011년 정보화통계조사[4]와 정보보호실태조사[1] 에서와 같이 모수 추정을 실시한다. 표본설계시 고려된 업종과 규모별 총 52개 층의 모집단 자료를 바탕으로 [표 6]의 '④가중치'를 산출한다.

본 연구의 모층계 추정식은 모집단이 L개의 층으로 구성되어 있다고 할 때, 각 층의 총계에 대한 추정량 (\hat{Y}_h)의 합계로 전체 모집단 총계 ($\hat{Y} = \sum \hat{Y}_h$)를 추정할 수 있으며, 여기에서 \hat{Y}_h 를 전수조사 부분과 표본조사 부분으로 구분하여 식(1)과 같이 $\hat{Y}_h = Y_{hc} + \hat{Y}_{hs}$ 로 추정한다.

$$\hat{Y}_h = \sum_{h=1}^L Y_{hc} + \sum_{h=1}^L \frac{N_{hs}}{n_{hs}} \sum_{k=1}^{n_{hs}} y_{hsk} \quad (1)$$

L : 층의 계수(업종 × 규모)

n_{hs} : h층의 표본 크기

y_{hsk} : h층 표본의 k번째 관찰값

$\frac{N_{hs}}{n_{hs}}$: 표본조사단위의 모집단 단위에 대한 가중값. h층에서

$\frac{n_{hs}}{n_{hs}}$: 응답률의 역수(응답률 100% - 표본추출률의 역수)

Y_{hc} : 전수조사에서 각층의 총계에 대한 추정량의 합계

모비율은 식(1) 모층계 추정식을 전체의 사례수로 나누어 주면 되므로 식(2)를 이용하여 추정하며, 모수 추정에 따른 오차는 95% 신뢰수준에서 ±2.79%이다.

$$\hat{p}_{st} = \frac{\hat{Y}}{N} = \frac{\sum_{h=1}^L Y_{hc} + \sum_{h=1}^L \frac{N_{hs}}{n_{hs}} \sum_{k=1}^{n_{hs}} y_{hsk}}{N} \quad (2)$$

III. 분석결과

수집된 처리방침 직접데이터 및 개인정보처리자 연관데이터에 대한 기초통계 분석은 엑셀(Excel) 2010과 SPSS 버전 18을 이용하여 분석한다. 조사된 표본데이터로부터 모수를 추정한 통계값을 이용하여 정보화통계조사[4][18][19] 및 정보보호실태조사[1][6] 결과와 비교한다. 그리고 본 논문에서 제안하는 방안의 타당성, 우수성 등을 제시할 수 있는 결과를 중심으로 분석하며, 개인정보보호 현황을 다양하게 보여줄 수 있는 홈페이지 연관데이터에 대한 분석결과는 제시하지 않는다.

3.1 처리방침 공시 현황

개인정보처리자가 처리방침을 공시하고 있는 현황은 개인정보보호법 인지 및 준수 현황을 살펴 볼 수 있는 중요한 지표중 하나로, 한국정보화진흥원에서 2003년부터 2011년까지 정보화통계의 항목으로 조사했으며[18], 한국인터넷진흥원에서도 정보보호 실태조사에 따라 2007년부터 2011년까지 '개인정보취급방침별 공개 여부' 조사결과를 통해, 간접적으로 파악할 수 있었다[1]. 그러나 [5]의 24페이지에 의하면 정보화통계조사에서는 2012년부터 개인정보보호법 제정에 따라 위반시 과태료 부과 관련 등 사업자가 준수해야 할 법적 의무사항을 포함한 민감항목으로 분류하고, 응답자 부담 경감 및 응답률 제고를 위해 조사 항목에서 제외 하였으며, 정보보호실태조사에서도 2012년부터 조사하지 않고 있어[6], 2012년 이후의 처리방침의 공시 현황을 알 수 있는 조사결과는 없는 상황이다.

본 연구에서 조사된 표본으로부터 추정한 전체 및 종사자 규모별 처리방침 공시 사례수는 [표 7]의 '① 사례수'이다. 이를 모집단 모수 69,837건과 종사자 규모별 모수로 나눈 '②비율'은 홈페이지를 운영하면서 개인정보를 수집하는 고용인수 5명 이상의 개인정보처리자중 처리방침을 공시하는 비율을 의미한다. 그리고 ③, ④는 [4]의 356페이지와 [19]의 336페이지에서 조사된 '처리방침 제정 및 운영 현황' 통계표에서 종사자수 규모에 해당하는 사례수를 더하고, 같은 방법으로 종사자수 규모별 홈페이지내 개인정보 수집 사업체 수로 나누어 산출 하였다.

이렇게 분석된 처리방침 공시 사례수는 총 52,760건으로, 이를 모수 69,837건으로 나눈 비율은 75.5%이다. 이는 2011년 6월에 조사[4]된 87.7% 보다는 12.2%가 낮으며, 2010년 6월 조사[19]된 75.3%보다는 0.2%가 높다. 종사자수 규모별로 살펴 보면, 250명 이상 규모의 개인정보처리자는 98.7%가 처리방침을 공시했으며, 2010년의 93.0%, 2011년의 98.4%보다 증가했다. 그러나 고용규모가 5~9명인 개인정보처리자의 경우 63.8%로 2010년의 76.5%, 2011년의 83.3%보다 낮다.

이는 2011년 9월 개인정보보호법 발효 이후 처리방침 미공개시 과태료 등의 규제가 강화된 이후 1년 6개월이 경과한 시점에서 조사된 결과라는 것을 고려할 때, 250명 이상 고용하고 있는 개인정보처리자의 경우 방침 공시율이 지속적으로 높아졌으나, 고용규모가

250명 미만인 개인정보처리자의 경우 2011년 정보화통계 조사의 결과보다 최대 19.5% 낮은 것으로 분석된다. 이는 조사대상 개인정보처리자가 해당 조사항목에 대해 실제로는 그렇지 않으면서도 처리방침을 제정 및 공개하고 있다고 답했을 가능성이 높은 것으로 분석되며, 2012년 정보화통계 조사부터 민감항목을 이유로 조사대상 항목에서 제외된 이유를 설명하는 것으로 분석된다. 본 연구에서 제안한 조사방법은 응답자의 부담을 고려하지 않고 처리방침의 공시 현황을 조사자가 직접 수집할 수 있는 방법이며, 설문조사를 통해 시간과 비용이 들어도 정확한 통계를 얻기 불가능한 상황에서 유일한 조사 방법으로, 향후 정보화통계 또는 정보보호 실태조사 등에서 본 연구에서와 같은 조사방법의 적용이 가능하다.

[표 7] 처리방침 제정 및 공시 비율 비교

구 분	제안방안		정보화통계조사		
	①사례수	②비율	③비율	④비율	
총합계	52,760	75.5%	87.7%	75.3%	
종사자수	1-4명	-	-	86.0%	67.7%
	5-9명	15,831	63.8%	83.3%	76.5%
	10-49명	25,157	79.0%	90.0%	80.4%
	50-249명	10,191	87.9%	97.3%	92.1%
	250명 이상	1,582	98.7%	98.4%	93.0%

3.2 처리방침 항목별 작성 현황

2011년도 정보보호 실태조사의 처리방침별 공개 항목 조사결과[1]와 본 연구에서 조사된 처리방침의 구성 항목별 작성 현황을 비교해 보니 전반적인 처리방침 내용의 충실도는 높아진 것으로 [표 8]과 같이 분석 된다. [표 8]의 구분에서 '기술', '미기술', '원칙만 기술', '세부내용 기술'은 [표 3]의 '기술Type' 인코딩 스킴에 따라 입력된 데이터를 기반으로 분석한 것으로, ②, ④, ⑦번 항목의 '세부 내용 기술'은 개인정보 제3자 제공을 받은 사업자 명칭, 위탁사업자의 명칭, 개인정보책임자의 이름 및 연락처가 구체적으로 기술된 경우, '원칙만 공개'는 처리방침 해당 항목에 대해 기술하고 있으나 구체적인 대상의 명칭을 기술하지 않은 경우를 의미한다. ①, ③, ⑤번 항목의 '기술'은 처리방침의 해당 항목의 내용이 기술되어 있는 경우를 의미하며, 모든 항목에서의 '미기술'은 해당 항목의 내용이 기술되어 있지 않은 경우를 의미한다. '세부

[표 8] 처리방침 구성 항목별 기술 현황

구 분	제안방안		2011년 정보보호 실태조사[1]
	사례수	비율①	
①개인정보 수집· 처리 목적	기술	49,693 94.2%	86.5%
	미기술	3,059 5.8%	-
②제3자 제공 목적, 제공 사업자, 기간	세부내용 기술	5,418 10.3%	57.9%
	원칙만 기술	34,609 65.8%	
③파기절차 및 방법	기술	32,014 60.9%	52.1%
	미기술	20,581 39.1%	-
④위탁업무 내용 및 수탁자	세부내용 기술	13,499 25.6%	42.8%
	원칙만 기술	17,832 33.9%	
	미기술	21,328 40.5%	
⑤권리의무 및 행사 방법	기술	35,874 68.2%	40.0%
	미기술	16,721 31.8%	-
⑦개인정보책임자 및 연락처	세부내용 기술	39,147 74.4%	45.6%
	원칙만 기술	4,735 9.0%	
	미기술	8,713 16.6%	

내용 기술'과 '원칙만 기술' 부분을 구분해서 분석한 이유는 해당 항목에 대해 단순 원칙만 기술한 처리방침과 세부 내용을 공시하고 있는 처리방침의 현황을 구분하여 처리방침 기술의 충실도를 좀 더 세부적으로 분석하기 위함이다.

그리고 [표 8]의 비율 ①은 공시된 처리 방침 중 해당 항목의 내용이 기술되어있는 비율을 의미하며, 처리방침을 제정 및 공시하고 있는 개인정보처리자의 수 52,760개로 해당 항목이 기술된 사례수를 나누어서 산출 한다.

2011년도 정보보호 실태조사에서 조사된 현황과 제안된 방안으로 조사된 결과를 비교해 보면, ①번 개인정보 수집/처리 목적 항목의 경우 86.5%에서 94.2%로, ⑦번 개인정보책임자 및 연락처 항목은 45.6%에서 74.4%로 크게 증가하는 등 전반적인 처리방침 내용의 충실도가 높아진 것으로 분석된다.

이는 2011년 개인정보보호법 시행과 함께 정부의 적극적인 홍보 및 계도에 힘입어 많은 개인정보처리자들이 처리방침을 보완한 것으로 판단되며, 2012년 정보보호 실태조사에서 '개인정보 취급방침별 공개 여부'에 대한 조사가 실시되지 않아[6], 본 연구에서 조사된 결과가 실제 처리방침의 내용을 기준으로 분석한 최근의 현황을 보여준다.

3.3 고유식별번호 수집 현황

정부는 정보통신망법 제23조 제2항에 따라 인터넷에서 주민등록번호의 사용을 2013년 2월 18일부터

제한하고 있다. 그러나 현재 주민등록번호 등 고유식별번호의 수집 및 이용과 관련된 통계는 부족한 상황이다. 본 연구에서 주민등록번호 등 고유식별번호의 수집 현황을 조사한 결과, 모집단 69,837개 개인정보 처리자 중 주민등록번호의 경우 39.51%인 27,596개가, 외국인등록번호, 여권번호, 운전면허번호의 경우 각각 0.96%, 1.04%, 1.49%인 672개, 726개, 1,039개 개인정보처리자가 수집 및 처리하는 것으로 조사 되었다.

[표 9]와 같이 업종별 고유식별번호 수집 현황 및 업종별 모수대비 비율인 ①을 살펴보면 주민등록번호는 모든 업종에서 수집하고 있으며, 업종별 수집 비율이 40%이상인면서 사례수가 500개가 넘는 업종으로는 제조업(45.5%, 3,593개), 도·소매업(41.4%, 2,992개), 숙박 및 음식점업(41.1%, 708개), 금융 및 보험업(48.9, 794개), 사업지원 서비스업(60.52%, 820개)이었다. 가장 많은 주민등록번호 수집 개인정보처리자가 있는 것으로 조사된 업종은 교육, 보건·복지, 예술·스포츠·여가업 등이 포함된 기타업종으로 13,637개 사례가 조사 되었다. 종사자수 규모별로는

[표 9] 업종별 고유식별번호 수집 현황

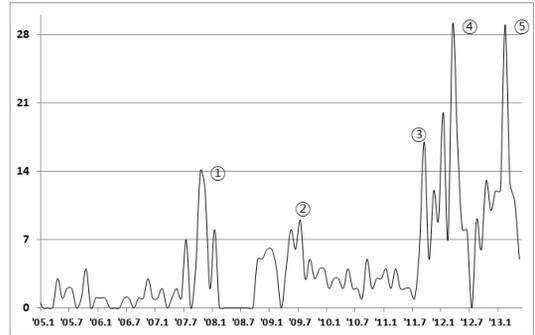
구 분	주민등록번호		외국인등록번호		여권번호		운전면허번호		
	사례수	①.%	사례수	①.%	사례수	①.%	사례수	①.%	
전 체	27,596	39.5	672	1.0	726	1.0	1,039	1.5	
1. 농림수산업	93	56.4	-	-	-	-	-	-	
2. 제조업	3,593	45.5	-	-	70	0.9	204	2.6	
3. 건설업	624	31.9	-	-	-	-	-	-	
4. 도매 및 소매업	2,992	41.4	8	0.1	8	0.1	358	4.9	
5. 운수업	167	22.8	-	-	62	8.5	-	-	
6. 숙박 및 음식점업	708	41.1	8	0.5	-	-	8	0.5	
7. 출판·영상·정보서비스업	864	25.1	171	5.0	157	4.6	62	1.8	
8. 금융 및 보험업	794	48.9	37	2.3	74	4.6	186	11.4	
9. 부동산 및 임대업	680	29.4	-	-	64	2.8	-	-	
10. 전문과학 기술서비스업	697	32.4	-	-	-	-	-	-	
11. 사설·사업지원 서비스업	820	60.5	52	3.8	72	5.3	-	-	
12. 협·단체 및 개인서비스업	1,929	38.6	208	4.2	69	1.4	65	1.3	
13. 기타	13,637	39.8	188	0.5	149	0.4	157	0.5	
종 사 자 수	5~9명	10,458	42.2	270	1.1	338	1.4	285	1.2
	10~49명	12,164	38.2	-	-	174	0.5	677	2.1
	50~249명	4,027	34.7	140	1.2	88	0.8	48	0.4
	250명 이상	946	59.1	262	16.4	126	7.9	28	1.8

250명 이상의 개인정보처리자가 59.1%로 수집비율이 높았으나, 10~49명 규모의 개인정보처리자가 12,164개로 가장 많은 수로 분석된다.

이러한 분석결과는 현재 정부에서 추진하고 있는 인터넷을 통한 주민등록번호 사용 제한 정책의 추진 성과를 평가하는데 있어 기준이 될 수 있는 자료이며, 정부에서 집중적으로 홍보, 기술 지원하기 위한 대상을 보다 구체적으로 식별할 수 있다는 점에서 의의가 있다.

3.4 처리방침 작성 시점 현황

연도별 처리방침 작성 시점을 분석해본 결과 [그림 2]에서와 같이 5회 정도 처리방침의 작성 빈도가 많은 시기가 있다. [표 10]과 같이 관련 법률의 제정 및 개정 시점과 비교 분석해 본 결과, 첫 번째 빈도가 높은 시점인 ①은 처리방침 미공개시 과태료를 부과하기로 한 정보통신망법 시행령(2007.7) 및 공공기관 처리방침의 인터넷 공개를 명시한 ‘공공기관 개인정보보호에관한법’ 개정(2007.11) 시점과 일치한다. 두 번째 빈도가 높은 시점인 ②는 처리방침을 공개할 때 ‘개인정보취급방침’ 명칭을 사용하게 한 정보통신망법 시행시기(2009.1)와 겹친다. 세 번째 빈도가 높은 시점인 ③은 개인정보보호법 및 시행령 제정 시기(2011.9)와, 네 번째 높은 시점인 ④는 개인정보보호법 본격 시행(2012.3.30)전의 시기와 일치하는 것으로 분석된다. 이는 처리방침의 공개에 영향을 미치는 관련 법 제·개정과 처리방침 작성 및 공개의 상관성이



(그림 2) 처리방침 작성 시점 현황

높다는 것을 의미한다. 이를 통해 처리방침이 아직까지는 자율적으로 지속 제·개정되기 보다는, 관련 법 제·개정, 정부의 정책 추진 등 외부의 영향요소에 영향을 받고 있다고 분석된다. 그러나 ⑤의 경우 2013년 초에 처리방침의 작성이 집중적으로 발생하였는데, 이는 처리방침 데이터의 맥락품질이 향상되고 있음을 의미하며, 지속적인 모니터링을 통해 개인정보보호 현황을 파악할 수 있는 가능성을 시사한다는 점에서 의미를 부여할 수 있다.

3.5 처리방침 명칭 사용 현황

처리방침의 명칭은 개인정보보호법에서는 ‘개인정보처리방침’으로, 정보통신망법에서는 ‘개인정보취급방침’으로 표시하는 것을 규정하고 있다. 개인정보처리자의 입장에서는 개인정보보호법과 정보통신망법의 적용에 대한 혼선을 빚고 있는데 처리방침 명칭 사용 현황 분석을 통해 개인정보처리자들의 법 적용 현황을 간접적으로 살펴볼 수 있다.

개인정보보호법이 제정된 2011년 3월 이후 작성된 처리방침의 명칭을 [표 11]과 같이 분석한 결과, ‘개인정보처리방침’이라는 명칭을 사용한 처리방침은 전체의 28.0%, ‘개인정보취급방침’이라는 명칭을 사용한 처리방침은 58.8%로 정보통신망법에 의한 ‘개인정보취급방침’이 더 많이 사용되고 있다. 특히 고용인수 5~9명인 개인정보처리자의 66.8%가 정보통신망법에 따른 ‘개인정보취급방침’이라는 명칭을 사용하고 있는데 반해, 고용인수 250명 이상의 개인정보처리자의 경우 개인정보보호법에 의한 명칭을 사용하는 비율이 60.3%로 정보통신망법에 의한 명칭 사용기관보다 많다. 일반법인 개인정보보호법 적용기관이 특별법인 정보통신망법 적용기관보다 많은 상황에서 정보통신망

[표 10] 처리방침 관련 법 제정 현황

법명	구분	시행일시	조문	주요내용
공공기관 개인정보 보호법	법	'07.5.17.개정 '07.11.18.시행	제7조(개인정보파일의 공고)	홈페이지에 공개 명시
			제7조의2(개인정보 보호방침)	관리책임관 및 개인정보보호에 관한 사항 포함 명시
정보통신망법	법	'07.1.26.개정 '07.7.27.시행	제27조의2(개인정보 취급방침의 공개)	개인정보취급방침 명칭 및 과태료 부과 최초 명시
	시행령	'09.1.28.일부 개정 및 시행	제14조(개인정보취급방침의 공개 방법)	구체적 명칭 표시 방법 최초 명시
개인정보 보호법	법	'11.3.29.제정 '12.3.30시행	제30조(개인정보 처리방침의 수립 및 공개)	개인정보처리방침 명칭 사용 및 과태료 부과 명시
		시행령	'11.9.29.제정 '12.3.30시행	제31조(개인정보 처리방침의 내용 및 공개방법 등)
	표준지침	'12.12.31	제7조(개인정보의 암호화)	고유식별번호, 비밀번호, 바이오정보에 대한 암호화 적용 완료

[표 11] 법 제정(2011.3월)이후 작성된 처리방침의 명칭 현황

구분	합계 (a)	개인정보 처리방침		개인정보 취급방침		기타	
		사례수 (b)	b/a (%)	사례수 (c)	c/a (%)	사례수 (d)	d/a (%)
합 계	15,060	4,219	28.0	8,852	58.8	1,990	13.2
5~9명	3,245	588	18.1	2,168	66.8	488	15.1
10~49명	7,389	2,061	27.9	4,491	60.8	838	11.3
50~249명	3,517	1,021	29.0	1,918	54.5	578	16.4
250명 이상	909	548	60.3	274	30.2	86	9.5

법에 따른 처리방침 명칭 사용 개인정보처리자가 많은 것은, 2007년 7월 처리방침 미공개시 과태료를 부과할 수 있도록 정보통신방법을 개정·시행 하면서, 많은 기업들이 정보통신방법에서 정한 '개인정보취급방침'이란 명칭을 우선적으로 사용하였기 때문으로 분석된다. 그러나 종사자수 250명 이상의 개인정보처리자의 경우, 해당 법률에 대한 검토 및 적용이 비교적 잘 이루어지기 때문에, 규모가 작은 개인정보처리자에 비해 개인정보보호법에서 정한 방침명칭을 사용하는 비율이 높은 것으로 분석된다.

IV. 결 론

홈페이지를 운영하면서 개인정보를 수집하는 개인정보처리자는 2011년 약 12만 6천개에 이른다[4]. 처리방침은 개인정보를 수집·처리하는 개인정보처리자라면 지속적으로 홈페이지 등에 공개해야 하며, 세부 내용에 대한 규정은 있으나 공개하는 형식에 대한 제한사항은 없다.

이러한 처리방침으로부터 일정 시간 간격을 가지고 수집한 처리방침 데이터는 개별 개인정보처리자의 개인정보 수집·이용·제공 및 보호 현황만 제공해 주는 것이 아니라, 우리나라 전체 개인정보처리자들의 법 준수 여부, 개인정보보호 현황 및 개인정보의 흐름(위탁, 제공) 현황에 대한 가치 있는 정보 원천으로, 대응량의 규모(volume), 다양한 형태(variety), 빠른 생성 속도(velocity) 등 빅데이터의 특성을 가진다[20].

본 논문에서 제시한 처리방침 데이터를 이용한 개인정보보호 현황 분석 방안은 ① 데이터 품질 4대 요소인 고유품질, 맥락품질, 접근품질 및 표현품질을 모두 만족하는 등 처리방침 데이터의 품질이 좋고, ② 기존 정보화통계 및 정보보호 실태조사 결과와 비교 분석해 본 결과, 처리방침 데이터를 이용한 분석이 의미 하며, ③ 다른 데이터와 결합하여 개인정보보호

현황에 대한 빅데이터 분석이 용이하여, 개인정보보호에 대한 실태 파악의 새로운 돌파구를 제시했다는 점에서 의의가 있다.

또한 연구수행을 위해 제시한 처리방침과 홈페이지 연관 데이터에 대한 메타데이터 항목 및 인코딩 스키마는 처리방침 정보를 보다 효율적으로 대규모로 수집하고 저장하기 위한 시스템 개발 및 다양하고 풍부한 데이터 분석을 위한 기반을 구축하는데 기여 가능하다.

그러나 본 논문에서는 수집한 처리방침의 수가 1,000개로 제한되어 통계적으로 작은 오차의 결과를 도출하기에는 한계가 있다. 분석을 위한 표본 수를 충분히 늘린다면 통계적으로 더 작은 오차의 분석결과를 산출 가능하다.

처리방침 데이터는 개인정보보호와 빅데이터 분야의 접목을 통해 우리나라 개인정보의 처리 및 보호 현황을 보다 더 정확하고 다양하게 분석할 수 있는 좋은 데이터 원천이다. 향후 다양한 통계적 분석 기법을 이용하여 처리방침 데이터와 개인정보보호 현황간의 인과관계를 파악하기 위한 다양한 연구가 기대된다.

참고문헌

- [1] 한국인터넷진흥원, 2011 정보보호실태조사 - 기업편, 2012년 3월.
- [2] USC Dornsife/Los Angeles Times poll <<http://dornsife.usc.edu/usc-lat-poll-privacy-march-2012/>>, Mar, 2012.
- [3] 김정덕, "개인정보보호를 위한 관리체제와 거버넌스," 정보보호학회지, 18(6), pp.1-5, 2008년 12월.
- [4] 한국정보화진흥원, 2011 정보화통계집, 2011년 10월.
- [5] 한국정보화진흥원, 2012 정보화통계집, 2012년 12월.
- [6] 한국인터넷진흥원, 2012 정보보호실태조사 - 기업편, 2013년 3월.
- [7] 이기현, 서의진, 개인정보 수집 제공 관리 실태 및 개선방안 조사결과, 거래조사 11(6), 한국소비자원, 2011년 9월.
- [8] 조선비즈, "통신사 개인정보, 보험·카드사 줄줄," <http://biz.chosun.com/site/data/html_dir/2013/03/11/2013031101674.html>, 2013년 3월 11일.
- [9] MK뉴스, "한 사이트에 '가입동의' 클릭 순간 - 내

- 정보 1000곳에 뿌려진다.” <<http://news.mk.co.kr/newsRead.php?year=2012&no=817369>>, 2012년 12월 9일.
- [10] W3C, “The Platform for Privacy Preferences 1.1 (P3P1.1) Specification,” 2006.
- [11] 노중혁, 진승현, “웹 환경에서 정책 기반 개인정보보호 기술”, 전자통신동향분석, 22(4), pp.144-155, 2007년 8월.
- [12] L. F. Cranor, S. Egelman, S. Sheng, A. M. McDonald and A. Chowdhury, “P3P deployment on websites,” *Electronic Commerce Research and Applications*, vol. 7, no. 3, pp.274-293. 2008.
- [13] S. Egelman, L. F. Cranor and A. Chowdhury. “An analysis of P3P-enabled web sites among top-20 search results,” *Proceedings of the 8th international conference on Electronic commerce: The new e-commerce: innovations for conquering current barriers, obstacles and limitations to conducting successful business on the internet*. ACM, pp.197-207, Aug. 2006.
- [14] Richard Y. Wang and Diane M. Strong, “Beyond Accuracy: What Data Quality Means to Data Consumers,” *Journal of Management Information Systems* vol. 12, no. 4 pp. 5-33, Spring, 1996.
- [15] ISO, “Information Technology - Metadata registries (MDR),” ISO/IEC 11179-1:2004(E).
- [16] 방송통신위원회, “개인정보취급방침의 전자적 표시 방법,” 방송통신위원회 고시 2008-4, 2008년 5월.
- [17] 대한상공회의소 기업정보 제공 시스템, <http://www.korchambiz.net/company/company_main.jsp>. 2013년 2월.
- [18] 한국정보화진흥원, “정보화 통계조사 통계자료”, <http://www.nia.or.kr/bbs/board_list.asp?nowpage=2&objpage=0&BoardID=201111281519062156&Order=010502&boardtype=&search_target=&keyword=&Allkeyword=>. 2009년 8월.
- [19] 한국정보화진흥원, 2010 정보화통계집, 2010년 10월.
- [20] Philip Russom, Big Data Analytics, TDWI Best Practices Report, TDWI, fourth quarter, 2011.

〈저자 소개〉



이 재 근 (Jae-geun Lee) 정회원

1992년 2월: 고려대학교 응용통계학과 졸업

2002년 8월: 순천향대학교 정보보호대학원 석사

2010년 2월: 순천향대학교 정보보호대학원 박사과정 수료

2008년 10월~2011년 2월 : 한국정보통신기술협회(TTA) RFID/USN 응용서비스 및 정보보호 프로젝트 그룹(PG3113) 의장

2003년 4월~현재: 한국정보화진흥원 개인정보보호기반부

<관심분야> 개인정보보호, RFID/USN 보안, 익명성 서비스, 빅데이터 보안



강 상 옥 (Sang-ug Kang) 정회원

1993년 2월: 경북대학교 전자공학과 졸업

1995년 12월: 미국 남가주대학교 전기공학과 석사

2011년 8월: 고려대학교 정보보안학과 박사

2002년 9월~현재: 한국정보화진흥원 개인정보보호기반부

<관심분야> 멀티미디어 보안, 개인정보보호, 융합 서비스



염 흥 열 (Heung-Youl Youm) 종신회원

1981년 2월: 한양대학교 전자공학과 학사 졸업

1983년 9월: 한양대학교 대학원 전자공학과 석사 졸업

1990년 2월: 한양대학교 전자공학과 박사 졸업

1982년 12월~1990년 9월: 한국전자통신연구소 선임연구원

1990년 9월~현재: 순천향대학교 정보보호학과 정교수

1997년 3월~2000년 3월: 순천향대학교 산학연권소사업센터 소장

1997년 3월~현재: 한국정보보호학회 총무이사, 학술이사, 교육이사, 논문지편집위원 위원장, 수석부회장(역), 학회장(역)

2005년~2008년: ITU-T SG17 Q9 Rapporteur(역)

2006년 11월~2009년 2월: 정보통신연구진흥원 정보보호전문위원

2009년 5월~현재: 국정원 암호검증위원회 위원

2009년~현재: ITU-T SG17 부의장/SG17 WP2 의장

<관심분야> 인터넷보안, USN 보안, IPTV 보안, 홈네트워크 보안, 암호 프로토콜