개인정보보호에서의 로그품질의 중요성
-보건복지 분야 사례를 중심으로-
The Importance of the Log Quality of Personal Information Protection
-A Case Study of the Health and Welfare Division-

이야리(한국보건사회연구원), 정영철(한국보건사회연구원), 김정숙(삼육대학교)

차 례

1. 서론

최근 개인정보 침해사고가 빈번하게 발생하고 있으며 정보보안 및 개인정보보호에 관한 관심도가 높아지고 있다. 기업들은 이윤 추구 및 서비스 제공의 편의상 명목을 위해 무분별하게 개인정보를 수집하고 이용에만 관심을 두고 수집한 개인정보의 관리에는 크게 신경 쓰지 않은 것이 사실이다. 실제 개인정보의 노출은 관리자 또는 정보주체의 무의식과 부주의로 인해 발생하는 경우가 많다고 한다. 홈페이지 게시판이나 댓글 등에 무의식적으로 올린 주민등록번호, 연락처, 신용카드번호, 은행 계좌번호, 이메일주소와 같은 정보가 악의적인 목적으로 사용되면 개인의 일상생활을 침해할 수 있을 뿐만 아니라 명의도용 등으로 이어져 신용카드 발급, 보이스피싱 등 경제적인 손실도 가져올 수 있다. 특히, 주민등록번호는 교육, 의료, 신용 등 개인의 민감한 신상에 관련된 정보와 연계될 수 있기 때문에 정보노출 시 개인의 손실뿐만 아니라 사회・경제적 손실까지 가져올 수 있다. 특히 환자들의 신상정보는 물론 건강정보까지 보유하고 있는 의료기관의 개인정보는 정보특성상 그 보호와 관리가 매우 중요하다.

정보가 기업이나 국가의 중요한 자산임은 굳이 언급할 필요가 없을 정도로 정보 그 자체가 가치라고 말할 수 있다. 이러한 정보를 구성하는 데이터에 있어서 유효성 여부는 곧 데이터 품질이라고 볼 수 있다. 제대로 된 데이터, 곧 데이터 품질이 좋은 데이터를 분석하고 활용해야 정보로써 가치를 가질 수 있기 때문이다[3]. 따라서 본 연구에서는 개인정보를 처리할 수 있도록 체계적으로 구성한 데이터베이스시스템(이하 "개인정보처리시스템"이라 한다)의 모든 기록을 담고 있는 로그(Log) 데이터에 관한 품질을 분석하여 문제점을 진단해보고 개인정보보호를 위한 로그품질의 중요성 제고 및 개선방안을 도출하고자 한다.

2. 배경연구

2011년 3월 「개인정보 보호법」이 제정되어 같은 해 9월부터 시행되면서 사회 각 분야에서의 개인정보보호에 대한 관심은 계속 증가하고 있다[1]. 더욱이 금융 및 유통분야에서는 개인정보 침해발생시 개인은 물론 국가나 사회적으로도 큰 경제적 피해를 가져올 수 있으며 특히 의료분야는 개인의 민감정보인 건강 및 의료정보를 다루고 있으므로 국가차원에서의 개인정보보호 수준향상에 대한 체계적 접근 및 노력이 요구된다. 2006년 OO기관 내부직원에 의해 가입자 개인정보 무단유출이 밝혀져 24명의 직원에 대한 징계가 발생하였고, 2008년 약국 직원에 의한 개인정보 72만 건 유출, 2010년 OO기관 내부직원에 의한 개인정보 10만 건 보유사실 적발 등 특히 공공기관 내부자의 개인정보 부정사용 및 오남용에 대한 관제의 필요성이 요구되었다.

2010년부터 보건복지부에서는 개인정보를 보유하고

있는 주요 정보시스템을 대상으로 개인정보 유출 및 오·남용 사전 예방을 위한 다양한 관제활동을 추진하였고, 2011년에는 '보건복지 개인정보통합관제시스템'을 고도화하여 보건복지부 주요 정보시스템의 개인정보 유출 및 오·남용 사고를 예방할 수 있도록 관제체계를 개선하였다(2012 연차보고서). 이후 현재까지 개인정보 관제대상 기관인 소속 및 산하기관들의 개인정보처리시스템에서 내부 업무에 대한 로그를 분석하여 개인정보의 목적 외 오·남용 및 유출 등을 사전에 방지하고자 적극적으로 노력해오고 있다[4][5].

개인정보 오·남용에 대한 일반적인 사례는 다음과 같다.

가. 업무 목적이외의 호기심에 의한 연예인이나 정치인 등 유명 인사나 동료직원, 그 밖의 지인에 대한 개인정보 처리(조회, 수정, 저장, 출력 등)

나. 업무 편의 등을 목적으로 공통 ID 사용한다거나 타인의 ID 도용 등 사용자 계정(아이디, 패스워드) 공유

다. 사용자별 혹은 업무별 부적합한 개인정보 접근 또는 관리자, 개발자 등에 대한 불필요한 개인정보 접근과 같은 접근권한 미준수

라. 퇴직자, 정보자, 휴직자 등에 의한 개인정보 접근과 같은 비권한자 접속 등이 있다.

3. 로그품질 현황분석에 따른 개선방안

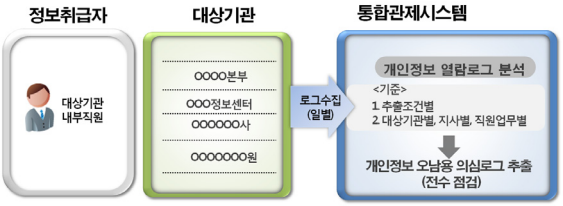
개인정보처리시스템을 통해 수집되는 업무로그는 5W1H(육하원칙)에 해당하는 표준화된 로그 칼럼으로 구성된다(표1 참조). 이는 업무상 개인정보 처리내역에 대한 책임 추적성 확보를 전제로 구성된 것으로써 고품질 로그가 확보됨에 따라 개인정보 오남용 의심사례 추출 조건의 유효성이 향상되고 궁극적인 목표인 개인정보 유출 등을 조기 발견 및 사전 예방하는 것이다. 그러나 표준화 로그의 품질 저하, 즉 데이터의 부재 또는 미진한 데이터로 구성된 로그로써는 원하는 추출조건에 의한 결과를 기대할 수 없다. 따라서 표준화 로그 품질개선에 의한 추출조건의 유효성을 향상시켜서 결과적으로 과탐, 미탐, 오탐 사례가 효과적으로 감소한다면 추출 업무의 효율성도 증가시킬 수 있다.

표준화 로그 수집 현황 및 개선점을 파악하기 위해 점검항목은 데이터 적재 시점에서의 오류 등으로 인한 로그 데이터 누락과 5W1H에 해당하는 특정 컬럼의 적합성 여부이다. 또한 점검 결과 개선 방안에 따른 계획을 수립하고 이행점검이 수반되어야 한다.

표 1. 로그품질 분석현황

번호	5W1H 기준 표준화 로그	표준화 로그 컬럼	표준화 업무로그의 필요성 및 분석 용도
1	WHEN	발생일/발생시간	사용자의 개인정보 사용 이력 분석 공휴일 사용, 비업무시간 내역 분석
2	WHO	사번/사원명/ 사용자ID	사원의 패턴을 분석하기 위한 주요 키값 공용 및 부적정한 사용자 정보 분석
3	WHERE	IP정보/부서정보/ 지사정보/소속정보	사용자 및 부서의 업무 발생시점 로그 분석 부서 및 지사 정보를 통해 권한 로그 분석
4	WHAT	고객주민번호/ 고객명/ 조회자료건수/ 조회건수/ 조회순번	사용자의 개인정보 사용 정보 분석 유의해야할 대상 유명인등 정보 분석 대량 유출 정보 확인 업무의 전후 순서를 분석
5	WHY	처리구분	업무 최종 결과 확인 분석
7	HOW	프로그램ID/ 프로그램명/ 요청URL	사용자가 개인정보를 사용하기 위한 프로그램 분석 사용자 및 부서의 프로그램 권한 분석

업무상 개인정보 오남용이 의심되는 로그를 추출하기 위한 일반적인 절차는 그림1과 같다[6].



▶▶ 그림 1. 오남용 의심로그 추출절차

시스템 사용자가 업무 목적외 개인정보 조회를 했는지 여부를 분석하기 위한 추출조건 사례 중에서 '주요인사 성명조회' 는 업무 열람자가 주요인사로 등록된 피 열람자의 개인정보를 수집하는 오남용에 대해 관련 로그를 추출하는 것이 목적이고 업무시스템 열람자가 주요인사의 성명이 들어 있는 프로그램을 조회 후 취득한 주민번호 등 정보를 이용하여 개인정보가 포함된 프로그램을 재차 조회한 경우에 해당한다(그림 2 참조).

▶▶ 그림 2. 추출조건 룰(주요인사 성명조회)

로그품질개선방안 마련을 위해 먼저 '13년 1월~2월 수집로스의 총합계 대비 적재된 업무로그 수에 대한 통계자료(예: 총 수집 로그 수 19,606,273건, 적재 업무 로그 수 19,478,102건)를 토대로 하여 계획, 점검 및 분석, 개선조치를 위한 일정협의, 이행 등 각 단계를 수립하여 '13.03.11~'13.06.30 동안 약 3개월간 실행하였다(표2 참조).

분석결과, 공통적으로 대상기관들에서는 표준화 로그의 일부 컬럼 데이터에 대한 미흡이 발견되었고, 일부기관들에서는 로그 적재 오류현상에 의한 누락부분이 발견되었다. 이에 따라 각 로그품질 현황 분석결과를 통보하고 문제점 개선을 위한 협의 및 조치 일정계획 수립이 마련되었으며 개선 결과에 대한 공유 등 적극적 노력이 수반되었다.

표 2. 로그품질 분석현황

기관명 (시스템)	협의일자	표준화로그 항목			
		계	정상	미흡	주요 품질 개선 항목
○○○	'13.03.26	22	17	5	- 적재로그 누락 - 육하원칙(5W1H) 컬럼 미흡
△△△	'13.03.26	22	9	13	- 적재로그 누락(컬럼 기준 위치 오류) - 표준화 로그 미흡
◇◇◇	'13.03.26	22	15	7	- 표준화 로그 미흡
□□□	'13.03.26	22	19	3	- 표준화 로그 미흡
♤♤♤	'13.03.29	22	17	5	- 표준화 로그 미흡
♡♡♡	'13.03.28	22	19	3	- 수집 업무 로그 파일명과 발생일 불일치 - 표준화 로그 미흡
♠♠♠	'13.03.28	22	10	12	- 표준화 로그 미흡
▽▽▽	'13.03.28	22	17	5	- 로그 적재 누락 - 표준화 로그 미흡

표2에서 △△△기관의 적재 오류 및 개선해야 할 사항을 정리해보면 표준화(업무로그) 수집 로그데이터 내의 사원코드(EMP_CD) 컬럼이 100% 정상적으로 채워져 있지 않았고, 업무 시스템 사용자가 내부 직원외의 외부 유관 단체 직원들도 존재하여 시스템 사용자 전체가 내부직원은 아니며 기관의 사원코드가 부여되지 않는 예외적 경우가 있었다. 이에 사원코드 부재로 인해 일부 로그 적재가 누락되는 문제가 발생하였다. 또한 수집된 업무로그에서 사원코드(필수정보)가 누락되어 수정이 필요하고 '관리자' 등 부적합한 사원명의 사용, 업무담당자의 소속 부서명의 누락, IP 누락, 조회 프로그램명 등 누락된 정보에 대해 데이터 추가가 필요하였다. 따라서 해당 기관에서는 정상서비스가 이루어지도록 개선조치계획을 수립하여 협의된 기간 내에 이행하도록 하였고 이후 6개월간은 상호 집중 점검기간을 거쳐 지속적인 품질수준을 유지할 것으로 기대한다.

각 대상기관과의 로그품질 개선을 위한 조치계획을 도출한 결과, 적재누락과 5W1H 로그 컬럼 미흡에 대한 두 가지 부분에 대해 개선 조치방안을 수립하였다. 먼저 로그적재누락 부분에 대해서는 수집 업무로그에서 사원코드(필수정보)의 누락에 대해 수정조치가 이루어져야 하고, 표준화 업무로그(업무이력) 내 주민번호는 암호화하여야 하며, 수집 업무로그 파일명과 파일 내의 업무로그 발생일이 일치하도록(예: 수집파일명≠업무발생일)한다.

그리고 표준화 업무 로그 내 특수 문자로 인해 기준 위치가 변경되는 오류 제거를 위해 특수문자가 입력되지 않도록 해야 한다. 로그 칼럼 미흡부분에서는 기본 육하원칙(5W1H)에 해당하는 표준화 업무로그를 구성하는 20여개 필수 칼럼 중 해당 기관별 미흡사항에 대해 추가 적재가 필요하다. 결과적으로 기관과 통합관제센터에서는 도출된 개선방안에 대한 조치계획에 따른 이행여부를 상호 점검할 수 있도록 상세 일정 및 협의가 수행될 것이다(표4 참조).

표 3. 로그품질 개선현황

연번	개선 항목	기관 조치 내역	점검 내용	
		계획	기간 ('13.06.03~6.14)	차주계획 ('13.06.17~6.28)
1	사번 누락 조치완료	정상서비스	점검 : 현재 점검결과 발생하지 않음 조치 : 개선조치 완료 결과 : 개선 조치 완료	적재로그 점검 미흡한 로그에 대해 수정요청
2	부적합한 사원명 사용 조치완료	정상서비스	점검 : 현재 점검결과 발생하지 않음 조치 : 개선조치 완료 결과 : 개선 조치 완료	추출로그 점검 미흡한 로그에 대해 수정요청
3	소속 부서명과 소속 지사명이 누락 조치완료	정상서비스	점검 : 현재 점검결과 발생하지 않음 조치 : 개선조치 완료 결과 : 개선 조치 완료	추출로그 점검 미흡한 로그에 대해 수정요청
4	사용자 IP 누락 수정조치중	정상서비스	점검 : 표준화 로그 테이블 점검 06.03~06.14 점검 5건 누락 조치 : 점검결과 통보 예정 (06.17) 결과 : 미완료	추출로그 점검 미흡한 로그에 대해 수정요청
5	프로그램 ID, 프로그램명, 요청 URL 누락 조치완료	정상서비스	점검 : 기관특성 확인하여 PRG_ID, REQ_URL 불가 단 PRG_NM 필수 조치 : 해당사항없음 결과 : 개선 조치 완료	추출로그 점검 미흡한 로그에 대해 수정요청
6	고객 주민번호와 고객명 누락 수정조치중	정상서비스 (5.21~)	점검 : 표준화 로그 테이블 점검 06.03~06.14 점검 주민번호 45,026건, 고객명 45,026건 누락 조치 : 점검결과 통보 예정 (06.17) 결과 : 미완료	업무로그 분석 누락 수정조치 요청 및 수신결과 미흡시 재전송

표 4. 로그품질개선 점검 결과

기관	점검 결과				이행 여부 점검
	로그 데이터 누락		필수컬럼 미흡		이행 여부
	개선 전	결과	개선 전	결과	
○○○			●	△	진행
△△△			●	△	진행
◇◇◇			●	△	진행
□□□			●	△	진행
♤♤♤			●	△	진행
♡♡♡	●		●	△	진행
♠♠♠	●	●	●	△	진행
▽▽▽	●	●	●	△	진행

※ ● -> 개선항목 있음 △ 미비(진행)재확인 필요

4. 결론 및 향후연구

본 연구의 로그품질개선 방안은 정착되는 단계이며, 한 단계 도약할 수 있는 과정에서 대상기관과의 최대한 협조가 절실하며 다음의 문제점들에 대해 계속해서 방안을 모색해나가야 한다. 첫째, 현재 수집 로그의 문제점은 변화되는 원본로그에 대응 방법의 부재이다. 사용자의 개인정보 접속 기록은 비정형 로그이며 표준화 로그는 관리 및 수정된 데이터이고 기존 비정형 로그에 대한 분석 방법이 현재로서는 없는 점이다. 최근 기술적 환경[7]에서는 정형/비정형 업무로그의 분석이 가능하지만 단순 시스템 변경만으로는 해결할 수 없으므로 표준화 로그를 구성하는 목적이 대상 기관에 유출 및 오남용 사례의 조기 발견 및 사전 예방을 위한 최선의 방안으로서 기관의 협조와 해결 방안을 공동 협의를 통해 도출해야 한다.

시스템 개발자의 관점에서는 예산이나 인력 등의 문제로 로그 품질 개선의 어려움을 호소하고 있으며 로그 수집 시스템의 변경이 발생할 경우 타 시스템이 연관되어 영향을 미치므로 원본로그 및 표준화 로그에 대한 우선순위는 시스템에 비해 높지 않다. 두 번째, 로그 생성 시스템의 차이로 인한 로그 품질 개선의 문제점이다. 기관마다 차이가 있으나 표준화 로그 생성은 API와 블랙박스 방식으로 크게 나눌 수 있다. API(Application Program Interface)를 통해 개인정보 데이터 접근시 로그를 생성하게 되면 표준화 로그 포맷에 맞출 수 있고 필요한 로그를 분석을 통하여 로그 생성 시스템에 업데이트도 가능한 대신, 운영 시스템의 업무를 모두 파악하고 있어야하는 개발팀이 필요하고 로그가 생성되는 시스템이 많은 경우 비용도 크게 발생하며 추가적인 Customizing 작업이 뒤따른다.

블랙박스를 이용한 패킷 분석 표준화 로그 수집에서는 Request, Result 등 네트워크 패킷을 통한 분석 솔루션 방식으로써 상대적으로 수집이 쉽고 적은 비용으로 높은 효율을 가져올 수 있다. 하지만 사용자 UI 화면에 대한 로그 정보 확보가 어려우며 표준화 로그 포맷에 맞추기도 어려워서 5W1H 필수 칼럼에 대한 별도의 Customizing 작업이 반드시 필요하다. 예를 들어 부서정보 및 사용자의 조회 정보, 조회수와 같은 데이터의 경우 패킷분석으로는 표준화 로그 생성이 불가하여 추가 작업이 필요하다. 즉, 실시간 배치 작업을 통한 표준화 업무로그의 생성이 불가능하여, 생성된 임시 로그 데이터에 대해 일배치 작업을 통하여 표준화 로그를 생성해야 한다. 마지막으로 다양한 시스템 환경으로 인한 기술적 이해 문제가 있다. 시스템을 위한 기반 개발 환경[8]은 기관의 환경에 따라 공공의 개발 프레임웍 및 Spring Framework, e-Gov Framework, Anyframe, X-Platform, MI-Platform, Servlet 등을 사용하고 있어서 대상 기관의 시스템 활용에 대해 각 담당자가 실제로 로그 생성을 위해서 필요한 기술적 이해의 수준이 넘어서는 상황에 있다. 따라서 표준화 업무로그의 품질을 개선하려는 개인정보 보호 담당자가 의지는 있으나, 기술적 방안을 제시하지 못하고 포기하는 경우도 많이 발생할 수 있다.

향후 로그품질 개선을 위한 연구계획으로는 기관이 생성하는 원본 로그에 대한 재분석을 통하여 표준화 업무로그의 품질 고도화 과정의 필요성 분석과 개선 및 개선조치 중의 통합관제 대상 기관의 표준화 업무로그를 유지하기 위하여 지속적 로그 품질 분석 등 모니터링이 수행되어야 한다. 그리고 현행 관리되고 있는 표준화 로그 중 기존 업무로그 이외의 대용량의 업무로그 수집 및 관리를 위한 업무 프로세스 설계 및 실시간 분석 및 분석결과에 대한 검토 방안을 구축할 예정이다.

참고문헌

[1] 개인정보 보호법령 및 지침・고시 해설, 행정안전부, 2011년 12월

[2] 국가정보보호백서, 방송통신위원회・행정안전부・지식경제부, 2011년 5월

[3] 데이터 품질관리 솔루션 자료집, 한국정보화진흥원, 2012년 11월

[4] 2012 개인정보보호 연차보고서, 개인정보보호위원회, 2012년 8월
[5] 보건복지부 개인정보보호 기본지침, 2012년
[6] 보건복지 개인정보 통합관제시스템 운영규정, 개인정보통합관제센터, 2012년 2월
[7] 장활식, 정대현, “컴퓨터 오남용의 의도와 행동을 결정하는 조직 및 개인적 특성”, 정보화정책 제20권 제1호 2013년 봄호
[8] 이주일외, “웹로그 분석을 위한 데이터 웨어하우스 시스템 구축”, 한국IT서비스학회, 2010년 춘계학술대회

저 자 소 개

● 이 야 리(Ya Ri Lee) 정회원

- 1990년 2월: 고려대학교 전자전산공학과 (공학사)
- 1999년 2월: 동국대학교 교육대학원(컴퓨터교육학석사)
- 2002년 8월: 동국대학교 컴퓨터공학과(공학박사)
- 2004년 3월 ~ 2012년 5월 : 대학 강사
- 2012년 6월 ~ 현재 : 한국보건사회연구원 초빙연구위원

<관심분야> : 개인정보보호, IT융합, 클라우드컴퓨팅, 모바일프로그래밍, 형식언어

● 정 영 철(Youngchul Chung)

- 1984년 2월: 연세대학교 가정학과(이학사)
- 1987년 2월 : 연세대학교 보건대학원(보건학석사)
- 1996년 8월 : KAIST 테크노경영대학원(경영정보학석사수료)
- 2003년 2월 : 가톨릭대학교 대학원(보건학박사수료)
- 2007년 4월 ~ 현재 : 한국보건사회연구원 연구위원

<관심분야> : 개인정보보호, 보건복지정보시스템 구축 및 평가, 건강정보서비스 제공, 복지정보서비스 제공

● 김 정 숙(Jung-Sook Kim) 정회원

- 1994년 2월: 광운대학교 전자계산학과(이학사)
- 1999년 2월: 동국대학교 컴퓨터공학과 (공학박사)
- 2000년 3월~2001년 2월: 김포대학 컴퓨터계열 교수
- 2001년 3월~현재: 삼육대학교 컴퓨터학부 교수

<관심분야> : IT 컨버전스, 임베디드시스템, 모바일 컴퓨팅, 웹프로그래밍, 프로그래밍언어, 컴파일러등.