정 보 보 호 학 회 지
제 30 권 제4호, 2020. 8

공공데이터 품질환경 내 데이터 오류의 발생원인별 보안기술 대응방안에 관한 연구

이 원 재*, 김 휘 강**

요 약

이 연구는 우리나라 정부의 공공데이터 공개 제도에 따른 공공데이터 품질관리체계를 이해하고, 공공기관이 신뢰성 있는 데이터를 위해 품질 점검을 시행하면서도 효과적인 관리를 하기 위한 방안에 관한 것이다. 공공데이터법과 공공데이터 품질관리체계를 이해하고, 저품질 공공데이터의 오류와 발생원인에 대해 알아본다. 오류 데이터 분석을 통한 보안위협에 따른 위험 분류를 통해 효과적인 대응방안을 도출하는 것을 목표로 한다.

이를 위해 공공데이터를 데이터 품질 점검하여 도메인별 오류데이터를 살펴보고, 오류데이터 발생원인에 대한 분석을 통해 보안위협과 공공데이터를 사용하는 사용자 측면과 기관 측면의 보안 문제를 분류하였다. 분류된 오류 발생원인별 보안문제를 기준으로 데이터 품질관리를 통한 개선방향을 제시하고, 품질관리 오류 개선방향별 데이터보안 정책별 보안기술을 비교 정리하여, 데이터 보안기술을 통한 품질관리 오류 개선 연계 대응방안을 제안하였다.

Ⅰ. 서 론

최근 정부에서는 공공데이터를 국민들에게 공개하면서 국가가 가지는 정보를 국민들이 사용할 수 있는 새로운 가치를 창출하기 위한 수단으로 활용하고 있으며 이를 제도적으로 지원하기 위해 많은 노력을 하고 있다.

그동안 관리와 활용이 어려웠던 공공정보, 즉 정부기관이나 공공기관이 생성하는 각종 데이터들을 소유하고 이용이 가능하게 되면서 막대한 양의 공공데이터는 개별 국민의 스마트 앱이나 웹을 통해서 제공되어 생활과 밀접한 연관을 가지고 될 뿐만아니라, 이러한 활용을 통해 국가의 새로운 성장 동력으로 재개발 되어지고 있다.

2013년 7월 30일 정부는 「공공데이터의 제공 및 이용 활성화에 관한 법률(이하 '공공데이터법'이라 함)」을 제정하여, 공공 데이터 제공 및 이용 활성화에 관한 사항, 국민의 공공데이터 이용권 보장, 공공데이터 활용을 통한 경제효과 창출, 다양한 기회를 이용한 민간시장 활성화를 목적으로 2013년 10월 31일부로 시행하고 있다.

정보화진흥원은 공공데이터의 중요성을 인식하고, 공공데이터의 민간 활용 활성화를 위한 품질제고 노력을 바탕으로 2014년 1월 공공데이터 품질관리 매뉴얼을 제정, 공공데이터 관련 법령 및 지침 개정, 품질관리 수준평가 제도를 시행하고 있다.

정부에서는 공공데이터 포털을 통해 정부 혹은 공공기관이 관리하는 공공데이터를 통합하고 국민에게 효과적으로 제공할 수 있는 통합 창구를 운영하고 있다.

현재까지 공공데이터 포털을 통해 민간에 공개되는 공공데이터 공개 현황은 [표 1]과 같다. 공공데이터 포털에서 제공되는 데이터를 참조하여 정리한 것이다.

포털에서는 모든 국민이 검색이나 편리한 분류 체계를 통해 원하는 데이터를 빠르고 정확하게 찾을 수 있도록 다양한 기능을 제공하며, 데이터 제공방식도 국민이 다양한 방식으로 공공데이터를 이용할 수 있도록

〔표 1〕 공공데이터 공개현황(공공데이터 포털, 2020)

(20.3.15기준)

공공데이터 유형	건수
파일데이터	30,646
오픈API	3,337
표준데이터	120
전체	34,103

* 고려대학교 사이버보안학과 (대학원생, hillsbe@korea.ac.kr)
** 고려대학교 정보보호대학원 (교수, cenda@korea.ac.kr)

XML, Json의 파일 제공형태와 Javascript, Ajax 등 오픈API 형태로 제공되고 있으며, 공공데이터 활용사례는 재난안전, 보건의료, 환경기상, 교육, 국토관리, 공공행정, 재정금융, 산업교용, 사회복지, 식품건강, 문화관광 등 다양하게 제공되어 활용되고 있다.

하지만 데이터 특성상 인터넷을 통한 전달 속도가 빠르고, 데이터의 신뢰성에 대한 문제가 가비지 데이터(Garbage Data) 등에 따른 잘못된 오류 데이터의 영향이 얼마나 심각한지는 다양한 매체를 통해 알려지고 있다. 특히 최근에 국민에게 필요한 공공데이터나 국가의 행정 데이터에 생긴 오류로 인해 이를 이용한 개인이나 조직에 서비스 신뢰도에 상당한 피해를 주었다는 소식이 빈번하다. 이뿐만 아니라 데이터 보안이 서비스의 보안에도 영향을 주는 문제점들이 발생하고 있다. 이러한 문제점들은 데이터를 소유하는 기관이나 데이터를 사용하는 사용자 모두에게 부정적인 위험을 줄 수 있다.

공공데이터 포털의 개발자 Q&A 기준으로 오류 건수는 총 1,979건, API 오류 건수는 5,173건 파싱, Ajax 등 스크립트 오류건수는 154건으로 집계되고 있다.

이 수치는 공공데이터 오류로 인한 직접적인 피해 사례는 아니지만 데이터 오류가 서비스의 오류를 통해 정보보안의 문제로도 확대될 수 있다는 개발자들의 우려가 제기 되고 있다.

현재 정부기관 및 지자체, 공공기관이 보유하고 있는 공공데이터의 오류율(5.19%)는 민간데이터의 오류율(2.10%)에 비해 상대적으로 높고, 데이터 품질에 대한 인식이나 품질관리체계가 부재하는 등 공공데이터 품질에 대한 관리적 · 운영적 문제가 다수 나타나고 있다[1].

공공데이터가 외부로 유통되거나 국민에게 공개 될 경우 국가 신뢰도에도 영향을 미칠 수 있어 그 파급효과는 민간보다 더 커질 수 있다. 또한 데이터를 생성 제공하는 공공기관에서도 이러한 공공데이터가 기관에 어떠한 영향을 미치는 지 부담스러운 실정이다. 이에 따라 공공데이터에 대한 중요성과 활용의 필요성은 커지고 있는 상황에서 정부는 보다 적극적으로 데이터의 신뢰도 유지를 위한 품질 점검에 대한 적용을 확대 시행하여 각 기관의 데이터 품질을 향상시켜나가고 있다. 하지만 데이터 소유자인 기관 입장에서 데이터 저품질의 원인인 오류데이터에 따른 보안위협 이나 위험에 관한 연구는 상대적으로 부족한 실정이다.

본 연구에서는 공공기관에서 관리되는 공공데이터의 품질을 점검하고, 점검된 실제 오류데이터를 기준으로 발생원인에 따른 위험요소를 도출하여, 위험별 대응방안을 보안기술을 통해 제시해보고자 한다.

Ⅱ. 이론적 고찰 및 관련 연구

2.1. 데이터 품질

데이터 품질관리에서 넓은 의미의 데이터는 조직의 전략과 목적을 달성하기 위하여 구축·운영되는 정보시스템과 관련된 모든 자료나 정보를 의미한다. 이때 데이터는 데이터베이스 내부에 저장되어 있는 데이터 값 이외에 데이터 모델이나 표준 데이터와 같은 구조 정보와 문서 형태의 산출물을 포함한다. 실제 품질 관리의 대상이 되는 좁은 의미의 데이터는 정보시스템에 저장된 디지털 데이터를 의미한다[2].

정부는 공공데이터의 제공 및 이용 활성화에 관한 법률 (약칭: 공공데이터법)을 통해 공공데이터, 품질관리, 점검 관련 근거를 마련하였고, 정부의 각 부처에서는 데이터 품질 관리 확산을 위한 추진계획과 점검가이드라인 등을 제시하면서 관리을 적극적으로 유도하고 있다.

데이터베이스진흥협회에서는「데이터 품질진단 절차 및 기법」(2009.6.24.)를 통해 국내의 데이터 품질 관련 연구에서의 품질지표를 값 · 구조 관점과 품질관리 프로세스 관점으로 분리하여 제시하고 있다.

데이터 관리 성숙도모형(2006)에서는 품질관리 프로세스 차원에서 ①정확성,②일관성,③유용성,④접근성,⑤적시성,⑥보안성 제시를 하였으나, 다만 아쉬운 점은 단일 데이터베이스 관점에서 데이터의 값 · 구조만을 염두에 둔 품질지표 정의임에 따라 국내 행정 및 공공기관이 당면하고 있는 문제인 데이터 보안, 오너십, 연계데이터의 일관성, 표준 준수에 대해서는 포괄하고 있지 못하고 있다.

행정자치부의 공공데이터 관리지침에서는 데이터 품질이란 데이터의 최신성, 정확성, 상호연계성 등을 확보하여 사용자에게 유용한 가치를 줄 수 있는 수준이며, 공공정보 품질관리 매뉴얼은 공공기관의 데이터 품질을 측정하는 기준인 8대 지표로 준비성, 완전성, 일관성, 정확성, 보안성, 적시성, 유용성과 지표별 세부 특성을 반영한 24개의 세부 지표를 제시한다[3].

공공데이터의 품질에 관한 초기 연구들도 대부분 정

확성만을 보는 미시적인 관점을 갖추고 있다. 그러나 실 사용에서의 문제나 활용에서의 영향은 정확성만으로는 품질을 보증하기가 어렵다. 더욱이 공공개방데이터는 민간에 개방되어 두루 다양한 사용자들이 다양한 요구를 가지고 활용되어지는 특성을 가지고 있다.

2.2. 데이터 품질 표준현황

국내외에서는 표 2와 같이 공공데이터 품질 관리를 위해 표준을 제정하고 관리를 적극적으로 유도하고 있다.

데이터품질 확보를 위해 국외에서는 ISO 9000(품질 경영 전반), ISO/IEC 12207, SPICE(소프트웨어 프로세스 생명주기), ISO/IEC 9126(프로세스 품질의 특성표), ISO/IEC 14598(제품 평가), ISO/TS 8000-150과 ISO/IEC 25012, DMBOK 가이드를 통해 공공데이터 관련 품질 표준을 지정하여 운영하고 있고,

국내표준으로는 한국데이터베이스진흥원에서 데이터베이스 품질[4]과 인증가이드(DQC-M, DQC-V, DQC-S)를 제시하고 있다.

DQC-M은 “행정 업무지원, 의사결정 정책지원, 지식의 활용 제공 등을 목적으로 운영되고 있는 정보시스템에 대한 데이터프로세스를 심사하여 인증하는 것”[5]을 의미하고 DQC-V는 “구축·활용 인 데이터베이스를 대상으로 도메인, 업무규칙을 기반으로 데이터 자체에 대한 품질향상요소를 심사·심의하여 인증하는 것”[6]을 의미하고 DQC-S는 “데이터베이스를 대상으로 DB 접근제어, DB 암호화, DB 작업결재, DB취약 분석 등 데이터베이스 보안에 대한 기술요소를 심사하여 인증하는 것”[7]을 의미한다.

〔표 2〕 데이터 품질 표준 현황

구분	표준	표준현황
국외	ISO 9000	품질 경영 전반
	ISO/IEC 12207, SPICE	소프트웨어 프로세스 생명주기
	ISO/IEC 9126	프로세스 품질의 특성표
	ISO/IEC 14598	제품 평가
	ISO TC184/SC4/WG13	공공데이터 관련 품질 표준
	ISO/TS 8000-150 ISO/IEC 25012 DMBOK 가이드	공공데이터 관련 품질 표준
국내	DQC-M	정보시스템 기반 데이터 프로세스
	DQC-V	데이터 자체 품질 영향요소
	DQC-S	데이터베이스 보안

특히, 정보보호 관점에서의 표준으로는 ISO/IEC 25012에서는 측정기준 #6 보안성(Security)에서는 권한 받은 사용자에게만 접근을 허용하는지 평가하는 것으로 평가했으며, 데이터 보안 인증(DQC-S)에서는 Audit 관점 데이터베이스 시스템에 보관되는 데이터를 허가받지 않은 외부의 침입으로부터 보호할 수 있는 체계가 마련되어있고 기술적으로 시스템보완을 수행하고 있는지에 대해 점검하고 인증한다[7].

2.3. 선행연구 분석

현재까지 발표된 데이터 품질 관련 논문으로는 크게 데이터 품질지표 및 품질점검활동 등 품질지표나 점검활동에 효율적인 관리절차와 점검방안에 기능적인 접근을 한 연구와 기업지배구조 내에서의 내부통제나 데이터 거버넌스 체계에 관한 정책적인 연구가 있었으며, 개인정보 등 민감정보에 대한 공공기업 법 준수를 위한 데이터베이스의 접근권한 통제 관련 개선안을 제시한 연구가 주를 이루고 있었으나, 데이터 저품질에 따른 위험이나 오류발생원인에 대한 위험요인 등 공공데이터를 소유한 기업관점에서의 위험요소나 분류체계, 대응방안에 대한 논문은 부족한 상황이다.

Kerr(2006), Batiniet al(2009), Madnick(2009)는 데이터 품질지표에 대한 논의가 초창기에는 정확성에서 시작되었으나 정보화 환경이 단일 데이터베이스에서 데이터베이스의 통합·연계로 진화하면서 데이터 값과 구조의 일관성, 보안 등 품질지표에 보다 많은 속성들이 추가되고 있다[8].

윤소영(2013) 는 데이터 관리절차와 관련하여 보안, 변환, 저장, 프로그램 개발 등과 같은 기능적 요구사항 기술을 제시하였다[9].

정승호(2013)는 공공기관의 데이터 품질에 영향을 미치는 요소로 데이터 보안,오너십, 연계데이터의 일관성, 표준준수를 표준지표로 제시하였다[10].

김현철(2015)은 비정형 데이터 중심의 공공데이터 개방에 따른 공공데이터의 품질과 관련된 특성들을 종

합적으로 정리하여 공공기관에게 필요한 데이터 품질 요인을 도출하고 공공데이터 개방 정책의 운영 방안을 제시하였다[11].

장경애,김자희,김우제(2015)은 데이터품질 표준(DQC-M)을 매개체로 RGT 기법을 사용하여 데이터품질 속성의 고객 인지구조를 도출하고, 도출된 Construct 간의 상관분석을 수행하여 AHP기법으로 평가속성의 가중치 및 우선순위를 선별하였다[12].

윤승영(2016)은 정보보안은 IT분야 뿐아니라 기업 거버넌스 차원에서 다뤄져야 하며, 기업 정보에 대한 보호를 위해 보안 분야에 대한 관리 과정에 안전장치가 제대로 작동하는 지에 대한 과정의 적절성도 중요성을 제시하였다[13].

장경애, 김우제(2017)는 기존의 IT 분야에서 벗어나 비즈니스와 연계하여 데이터 거버넌스 측면에서 관리되어야 한다고 하였으며, 데이터 거버넌스 속성지표를 데이터 통제영역, 품질영역, 조직영역으로 나누어 각 영역별 속성지표의 가중치와 우선순위를 선별하였다[14].

현재까지의 연구는 선행논문 이외에도 데이터 품질 환경하에서의 보안관련 점검지표 및 연구보고서가 품질 체계구축 및 품질점검 방안에 대한 지표를 제시하고 있다. 이렇듯 선행연구는 품질관점에서의 정보화 환경을 제시하고, 품질관리의 요구사항 기술, 관리 절차, 평가기준으로 보안 기능을 제시하고 있으며, 기업 내부통제 관점에서의 데이터 관리 상 보안 이슈에 대한 적절성, 데이터 거버넌스 내 영역별 속성으로 보안을 제시하고 있었다.

이에 본 연구에서는 기존의 지표들 내에서 의미있는 속성지표들을 선별하여 속성지표 점검기준을 도출한 뒤, 공공데이터 샘플데이터를 테스트데이터 셋으로 구성하여 도출된 점검기준에 따라 품질점검을 실시하도록 하겠다. 이후 오류 데이터 위험 분석을 통해 공공데이터 오류데이터에 대한 상호관계 분석을 통해 데이터 위험에 따라 어떻게 대응할 수 있는 지에 대해서 방안을 모색해 보도록 하겠다.

Ⅲ. 품질관리 수준 평가 분석

3.1. 품질관리 수준 평가 모델에 따른 실데이터 분석

3.1.1. 데이터 품질 분석 개요

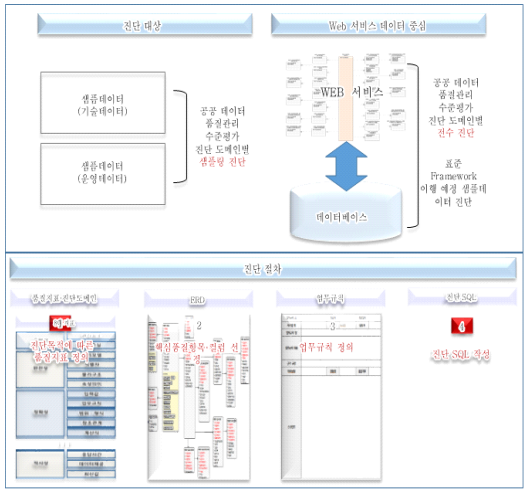
데이터 품질 진단대상과 진단절차는 [그림1]과 같이 공공데이터 품질관리 수준평가 진단 도메인별 샘플데이터를 진단대상으로 하며[16], 현재 웹 서비스되고 있는 데이터를 진단 대상으로 하였다. 진단절차는 품질지표, 대상선정, 진단규칙을 정의하고, 도구 및 프로파일링을 거쳐 오류 원인을 분석하였다.

데이터 품질 진단절차는 ① 진단목적에 따른 품질지표를 정의하고, ② 핵심 품질 항목 컬럼을 선정하여, ③ 업무규칙 정의하여 ④ 품질 진단 SQL 작성하는 순으로 진행하였다.

① 진단목적에 따른 품질지표는 아래 [표 3]과 같이 공공 데이터 품질관리 수준평가 추진계획의 오류 데이터 관리 평가항목 충족조건을 유효성, 정합성 부문의 총 5가지 진단항목을 기준으로 수행하였다.

② 핵심 품질 항목 컬럼은 각 품질지표의 진단항목별로 진단속성을 구분할 수 있는 [표 4]과 같이 진단컬럼을 선정하였고, 진단속성에 따라 ③ 업무규칙 정의하여 ④ 품질 진단 SQL 작성하여 분석하였다.

진단대상은 공공데이터 특성에 따라서 기술데이터와 행정데이터를 포함하여 86개 테이블을 선정하여 관련



〔그림 1〕 데이터품질 진단대상 및 진단절차

〔표 3〕 진단목적에 따른 품질지표 정의〔16〕

구분	진단항목	설 명
유효성	여부	'여부' 데이터(oo유무 등) 유효 값 오류 진단
	날짜	날짜 형식의 '날짜' 데이터 유효 값 오류 진단
	코드	'코드' 데이터 유효 값 오류 진단
	번호	규칙이 있는 '번호' 데이터 유효 값 오류 진단
정합성	참조값	관계를 갖는 데이터 사이의 일관성(참조무결성) 오류 진단

〔표 4〕 핵심 품질 항목 컬럼 선정〔16〕

진단항목	진단속성	진단컬럼	업무규칙
여부	NULL, 'Y'/'N', True/False	필요여브, 삭제여브	여부 적재규칙
날짜	'YYYY-MM-DD', 윤달요건, 선후행 관계	작업시간, 계약기간	년-월-일 데이터규칙
코드	코드 값 범위, 유효 여부 확인	부서코드, 관계코드	코드값 범위 규칙
번호	요청번호 범위, 번호길이	허가요청 번호, 사번	번호형식 규칙
참조값	후행일자/선행일자, 논리정합성	조치시작일시/종료일시	시간순서 규칙

지표에 따라 101개 컬럼을 분석 대상으로 하였다. 진단 대상 테이블의 전체 컬럼 중 5개 진단항목에 해당하는 컬럼의 전체 데이터로 진단 도메인별로 각 도메인에 해당하는 컬럼을 가진 진단대상을 샘플링하였다. 샘플링한 테이블들은 테스트 데이터 셋으로 구분하여 진행하였고, 분석방법은 SQL 쿼리를 이용한 수동 분석 작업과 진단규칙을 적용한 자동화 도구를 이용한 분석으로 진행하였고, 진행 중 발견되는 이슈가 될 만한 부분은 수동작업을 통해 심층적으로 진단하였다.

주요 분석 SQL 쿼리에 대한 오류 집계 쿼리는 품질 진단항목에 대한 정의된 업무규칙에 따라 [표 5]와 같이 각각 작성되어 분석하였다.

〔표 5〕 품질 진단항목 별 SQL 쿼리 예시

진단항목	업무규칙	품질 진단항목 별 SQL 쿼리 예시
여부	여부 적재규칙	select /*+ rule */ a.table_name, b.comments entity_name, c.column_id column_position, c.column_name, d.comments attribute_name, c.nullable from all_tables a, all_tab_comments b, all_tab_columns c where a.owner = '진단DB' and b.owner = '진단DB' and b.table_type = 'TABLE' and c.owner = '진단DB' and c.table_name = a.table_name
날짜	년-월-일 데이터 규칙	select 품질지표, 진단도메인, '전체', 데이터명, null 엔터티명, sum(decode(진단구분, '전체집계', 집계건수, 0)) 전체건수, sum(decode(진단구분, '오류집계', 집계건수, 0)) 오류건수, round(sum(decode(진단구분, '오류집계', 집계건수, 0))/sum(decode(진단구분, '전체집계', 집계건수, 0))*100, 5) 오류율 from isp_분석결과 where 진단구분 in ('전체집계', '오류집계') and 품질지표 = '유효 and 진단도메인 = '날짜' group by 품질지표, 진단도메인
코드	코드값 범위 규칙	select to_char(sysdate, 'yyyymmdd') 진단일자, 업무구분, 품질지표, 진단도메인, 진단테이블명, 진단엔티티명, 진단컬럼명, 진단속성명, 참조테이블명, 진단구분, 집계건수, 사례데이터 from (select '작업' 업무구분, '유효성' 품질지표, ' 금액 ' 진단도메인, '오류집계' 대상테이블명, ' 진단테이블명, '작업시간' 진단엔티티명, ' wprice ' 진단컬럼명, ' 금액 ' 진단속성명, null 참조테이블명, '오류사례' 진단구분, null 집계건수, to_char(수당, ' 999999999999999 ') 사례데이터 from DB명.적업테이블a where 수당is null or 수당< 0 group by 수당) where rownum<= 10
번호	번호형식규칙	select to_char(sysdate, 'yyyymmdd') 진단일자, 업무구분, '유효성' 품질지표, '번호형식' 진단도메인, 진단테이블명, 진단엔티티명, 진단컬럼명, 진단속성명, null 참조테이블명, '전체' 진단구분, count(*) 집계건수, null 사례데이터 from 진단DB.진단table
참조값	시간 순서 규칙	select '공사작업' 업무구분, '정확성' 품질지표, '참조무결성' 진단도메인, 진단테이블명, '공사도구' 진단엔티티명, '작업번호' 진단컬럼명, '작업no' 진단속성명, 참조테이블명, '오류집계' 진단구분, count(*) 집계건수, null 사례데이터 from 진단DB.진단table a where not exists (select 1 from 진단DB.참조table x where x.작업no = a.작업no)

3.1.2. 데이터 분석 결과

데이터 품질 측정은 공공데이터 품질관리 수준 평가 모델에 따라 진행하였으며 업무적으로 활용하는 데이터 모델, 각종 서식 등의 코드를 수집하여 데이터 표준화 지침의 유사성 분석 내용을 적용하여 유사한 코드를 군집화하였으며 코드군 내 영향도가 가장 높은 코드를 기준으로 측정을 진행하였다. 도메인별 값 진단 결과는 [표 6]과 같다.

데이터 분석 결과 과거 누적분 데이터에서 각 진단 도메인별 진단규칙에 어긋나는 데이터가 집중 발견되었고, 특히 여부도메인, 율도메인, 패턴도메인, 코드도메인 순으로 오류 건수가 발견되었다. 오류율이 높은 도메인에 대한 심층 분석을 통하여 기준데이터가 클수록 데이터 중복이 많이 발견되었다.

데이터 품질 점검을 통해 추출한 오류데이터를 기준으로 분류를 해보면 크게 ① Null 오류(여부, 율 도메인), ② 형식 오류(패턴, 금액, 수량, 코드, 관, 날짜 도메인), ③ Rule 위반(선후관계, 컬럼간 일관성 도메인)의 세가지 분류로 구분할 수 있다.

〔표 6〕 도메인별 값 진단 결과

분석대상	분석대상	테이블	오류건수	오류율(%)
도메인	여부도메인	5	4,223,264	69.9523
	수량도메인	5	19	0.0075
	금액도메인	2	20	0.0432
	율도메인	5	10,221	18.1923
	날짜도메인	5	31,877	1.0332
	패턴도메인	5	154,184	15.4871
	코드도메인	10	66,205	7.3688
	관도메인	39	2,082	0.013
데이터 규칙	선후관계 일관성	5	393,503	13.4293
	컬럼간 일관성	5	622	8.8077
합계		86	4,881,997	16.1306

분류별 오류데이터의 특징을 살펴보면 [표7]과 같이 정리할 수 있다.

오류데이터 분류를 기준으로 데이터 오류데이터가 공공기관 측면에서 어떤 원인으로 발생을 하며, 발생원인에 따라 영향을 줄 수 있는 위험요소에 대해 구분해 보도록 하겠다.

〔표 7〕 분류별 오류데이터 특징

도메인	오류데이터 구분	특징
여부, 율	Null 오류	이관 데이터, 컬럼의 추가 변경 등 관리 오류
패턴, 금액, 수량, 코드, 관, 날짜	형식오류, 범위 오류	응용시스템 통제 미흡, 사용자 실수
선후관계, 컬럼간 일관성	Rule 위반	데이터 지침 제도 부재, 시스템 변경관리 컬럼간 일관성 검증 부재

3.2. 데이터 품질 오류 실증분석

3.2.1. 데이터 품질 오류발생원인

오류 발생 분석을 위해 한국데이터베이스진흥원의 데이터품질 가이드라인에서 제공하는 주요 오류발생원인을 기준으로 주요 오류 데이터 발생원인에 대해서 데이터 입력측면에서는 이용자 혹은 데이터 입력자에 대한 입력 오류가 주로 있다고 보았고, 특히 이 부분은 입력 응용시스템의 수준에 따라 오류율이 달라 질 수 있다.

데이터 흐름측면에서는 시스템 통합이나 데이터 이관, 추출, 변환 등 대부분 시스템 관리 측면에서의 주로 데이터 관리자를 통한 오류로서, 관리자의 기술이나 지식 수준에 따라 오류율이 달라 질 수 있다.

데이터 재검증 활동 측면에서는 데이터 관리 체계, 즉 관리 프로세스에 대해서 재검증, 변경 관리에 따라 오류율이 달라 질 수 있다.

지금까지 추출된 오류 데이터를 기준으로 주요 오류 데이터 발생원인을 오류별로 구분을 위해서는 대상 시스템의 담당자 및 관리자 7인과의 협의를 통해 오류데이터에 따른 발생원인 분류에 대한 의견을 수렴하였고, 최종적으로 연구자가 관련자료 및 오류 데이터 수기분류 결과를 종합하여 [표 8]과 같이 오류유형에 따른 주요 발생원인을 분류하였다

분류된 오류유형에 따른 주요 발생원인 결과를 살펴보면 크게 보면 데이터표준, 데이터구조, 데이터값으로

〔표 8〕 오류유형에 따른 주요 발생원인

오류 유형	주요 오류	주요 발생원인
데이터 표준	사전 정의된 규칙 위반	표준용어, 도메인, 사전정의 규칙 미준수
	코드 형식 오류, 코드 정의 위반	규칙 변경, 컬럼 추가 등 시스템 관리 부재
	선·후행 관계, 참조데이터 오류	참조키(Foreign key) 관계 설정, 트리거 제약 프로그램 기능 미흡
데이터 구조	데이터모델 현행화 미흡	미사용 컬럼, 식별자 미정의
	데이터 관리 체계 부재	번호 형식 오류, 범위 오류
	데이터 식별 정보 오류	기본키 등 식별자 미정의
	미사용 테이블, 데이터 중복	미사용 컬럼, 동일 구조 테이블 중복
데이터 값	초기데이터 입력 오류	입력자 오류, 경험부족, 교육부족
	응용프로그램(처리,검증) 미흡	필수 입력값, 코드관리 값 검증 미흡

구분지어 살펴볼 수 있다. 주요 오류는 주로 입력자에 의한 오류, 응용프로그램의 검증의 직접적인 원인과 데이터 모델, 관리 체계 등의 부재, 데이터 표준 부재등의 간접적인 원인으로 구분 할 수 있다.

3.2.2. 공공데이터 오류발생원인에 따른 보안위협요소

데이터저품질에 따른 직·간접적인 오류 원인에 대한 보안 위협 사례로 악의적인 의도를 가진 사람들이 공공데이터를 사용하여 오류데이터에 대한 부적절한 분석결과를 통해 사실을 감추거나 잘못된 정보로 범죄에 악용하거나, 악의를 가지고 있는 내부 관계자에 의한 데이터 수정 변경에 따라 기업 정보를 훼손 혹은 유출시킬수 있는 데이터 보안 문제가 발생할 수 있다. 시스템적으로는 데이터 이전 과정에서 오류가 나타나면서 몇 시간 동안 서비스가 중단되거나, 데이터 저품질에 대한 공개되지 않은 위협으로 오류 데이터를 사용한 악의적인 행위에 대한 보안 위협요소는 항상 존재한다.

이러한 경우 공공데이터법 제 28조에 3항(제공중단 사유)에 따라 3. 공공데이터를 범죄 등의 불법행위에 악용하는 경우는 공공데이터 제공을 중단하고 있지만 모든 공공데이터 활용 사례를 알기는 어려우므로 공공데이터의 소유기관에서는 데이터에 대한 품질을 유지하고 공공데이터에 대한 보안위협요소에 대해 대응을 해야 할 것이다.

공공데이터로 인한 보안 이슈는 데이터를 사용하는 개인 사용자 영역과 데이터를 소유하는 기관 데이터 관리자 영역으로 나눌 수 있다. 개인 사용자는 공공데이터를 실제로 이용하여 정보를 생산해 내는 입장이며, 공공데이터의 신뢰도가 생산해 내는 정보의 신뢰도로 이어지는 특징을 가지고 있다. 보통 초반에 언급한 국가 공공데이터 포털을 통해 파일데이터, 오픈API, 표준데이터 등의 형태로 데이터를 제공받고 있으며, 서비스하는 방식은 주로 웹을 기반한 서비스였다.

공공데이터를 사용하는 사용자 관점에서 나타날 수 있는 보안 문제는 다음과 같다.

- 오류 데이터로 인한 웹 서비스 오류 및 장애
- 공공 데이터에 대한 상업적 목적의 신뢰성

공공데이터를 제공하는 기관의 입장에서는 공공데이터에 대해 비밀 정보의 노출 없이 데이터의 품질이 보장되면서 신뢰성이 유지되기를 원한다. 기관 입장에서 데이터 저품질의 보안 문제는 아래와 같다

- 민감 정보 노출(기업비밀 정보)
- 운영시스템 오류/중단/장애
- 데이터 손상 및 정보 유출
- 법/규제 미준수에 따른 기업 이미지 하락

위와 같은 보안위협요소에 대해 오류 유형을 기준으로 발생원인별 위협요소별 기관 담당자 관점에서 위협요소 및 보안 문제를 정리하였다.

데이터표준은 기관 내 데이터 정책의 수준에 따라 달라지며, 법/규제 준수를 위한 표준화 제도 등 이를 실행할 전문인력이나 기술 확보를 통해 대응할 수 있다.

데이터구조는 응용프로그램이나 데이터 저장소에 대한 트랜잭션처리 미비가 발생원인으로 분석결과나 신뢰성, 오류 데이터에 따른 시스템 오류가 위협요소이며 이는 데이터 신뢰성에 대한 분석결과 활용이나 응용시스템 자체 보안 문제로 인식되었다.

데이터값은 이용자에 대한 경험이나 실수에 의한 것으로 잘못된 정보 제공, 데이터 관련 기술 오류, 데이터 분석 조작으로 기업정보를 훼손 하거나 유출 시킬 수 있는 보안 문제를 유발시킬 수 있다.

각 오류유형별로 발생원인에 따른 위협요소를 종합적으로 정리하면 [표 9]와 같다.

〔표 9〕 데이터 저품질 주요 오류데이터 종류 및 위협 요소

주요 발생원인	위협요소	보안 문제
표준화 수준 미비 (표준코드, 도메인)	표준화 기준 수립 전문 인력 부족 분석 기술 부족	법/규제 미준수
주요 업무규칙 미처리 입력절차의 문제점 입력프로그램 오류 트랜잭션처리 미비	부적절한 분석결과 원천 데이터 신뢰성 불필요한 데이터 시스템적 오류	공공 데이터 신뢰성 서비스 오류(장애) 응용프로그램 보안
이용자 입력오류 시점차이 입력전 자료이관 입력통제미비 유효성 체크로직 미비	잘못된 정보 제공 관련 기술적 오류 데이터 분석 조작	기업 정보 훼손 및 유출 응용프로그램 보안

데이터표준 유형의 제도적 발생원인의 경우 공공데이터별 성격에 따른 표준화 기준 수립을 해야 하므로 공공데이터 점검 기준에 따른 각 기관의 표준화 기준 수립 문제가 있을 수 있으며, 공공데이터 개인정보유출이나 프라이버시 침해 문제, 데이터 정책의 전문 인력 육성이나 분석 기술이 부족으로 사람의 인식이나 능력의 부족이 초래하는 위협이 존재한다.

데이터구조 유형의 업무규칙이나 입력절차의 문제로 인한 입력프로그램의 오류는 오류바이러스나 웜 또는 DDOS 등 해킹에 악용될 수 있는 소지가 있으며, 시스템 데이터 훼손이나 손실의 위험이 있다.

데이커값 유형의 데이터 입력자 경험 격차 문제로 인한 데이터 값의 오류에 대한 문제는 주요 분석대상 데이터 개인의 행위기록, 패턴, 율 등 신뢰성 낮은 비정형 데이터이고 사전 정의 되지 않았을 경우 분석 오류에 대한 위험으로 이어질 수 있다.

데이터표준에 대한 위협요소는 전문인력 확보나 분석기술 확보 등 기관의 정책적 대응방안을 통해서 가능하며, 데이터구조나 데이터값에 대한 위협은 인적 보안 측면에서는 내부인력에 의해 의도된 위협에 대해, 내부자 보안 인식 교육을 통해 장기간 기관의 정책적 방안을 통해 대응이 될 것이며, 기술적 보안측면에서는 오류유형에 따른 품질측면에서의 대응방안과 데이터 관련 보안 정책별 보안 기술에 따른 대응방안이 있다. 이에 대해 저품질 데이터를 보호하기 위한 효과적인 기술적 대응방안을 제안하도록 하겠다.

Ⅳ. 데이터 품질 오류 발생원인에 대한 대응방안

4.1. 데이터 오류 개선방향 및 정책별 보안기술

앞서 살펴본 오류유형별 해당 위협요소에 대한 보안 문제에 대해 공공데이터 품질관리 측면에서 품질 오류 개선방안에 대해 알아보겠다. 한국정보화진흥원의 공공데이터 품질관리 매뉴얼[16]에서는 제시하는 오류 발생원인에 대한 개선방향을 [표 10]과 같이 주요 발생원인에 대해서 정리하였다.

이에 대해 살펴보면, 정보보안정책에 따른 데이터 관련 보안 정책별 보안 기술과 비교하여 중복되는 기능이나 역할에 대해 주목할 필요가 있다. 이에 보안 정책별 보안 기술과 기능 비교를 통해 효과적인 오류 발생원인

〔표 10〕 데이터 품질관리를 통한 오류 발생원인 개선방향

주요 발생원인	개선방향	설명
표준화 수준 미비 (표준코드, 도메인)	업무절차 개선	데이터 표준 수립 및 관리 정책 마련
주요 업무규칙 미처리	업무규칙 검증 응용프로그램 개선	데이터 모델 속성 정의 및 데이터 검증로직 적용 및 보완
입력절차의 문제점	오류 데이터 원인 분석 응용프로그램 개선	데이터 모델 재설계 및 변경 입력단계 데이터 검증
입력프로그램 오류		
이용자 입력오류	사용자 가이드 업무절차 개선 응용프로그램개선	사용자 가이드 제공 오류추정 데이터 분석 및 정비 데이터 검증로직 적용 및 보완
입력자동기 부여미비		
발생과입력시점간 차이		
입력전 자료이관	과거 데이터 정합성 검증	규정에 따른 관리항목 추가 데이터 이력관리
입력통제미비	응용프로그램 개선	데이터 검증로직 적용 및 보완
유효성 체크로직 미비		
사용자 인증 오류	응용프로그램 개선	오류 프로그램 권한 재설정

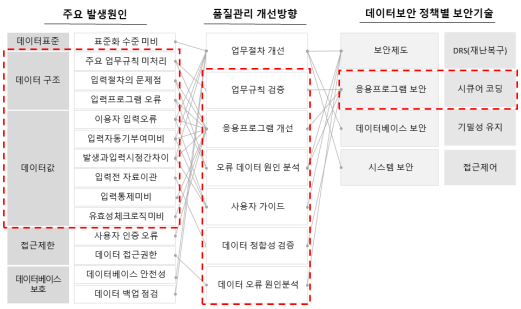
에 대한 대응방안을 제안하고자 한다. 현재 공공기관에서 운영하는 일반적인 정보보안정책에 따른 데이터 관련 보안 정책별 보안 기술을 열거하면 [표 11]과 같다.

〔표 11〕 데이터 관련 보안 정책별 보안 기술

보안정책	보안기술	설명
보안 제도	DRS[1] (Disaster Recovery System)	· BIA (Business Impact Analysis) · 백업센터 운영형태/기술형태/위치 · Hot Standby, DB shadowing, 원격 백업, Log) · 데이터 중요도, 우선순위
데이터베이스 보안	기밀성 유지	· 체크섬(Checksum), 해싱(Hashing)

보안정책	보안기술	설명
		· PKI(Public Key Infrastructure)
	태깅(Tagging)	· 데이터에 부가정보를 추가하여 각 데이터 간 접근제어를 가능하게 하는 작업
	접근제어	· 데이터 R&R을 위한 DAC, MAC, RBAC 구성
침해 대응	표준화 수준 미비 (표준코드, 도메인)	· 표준화 기준 수립 · 정보 이용범위 불명시 · 전문 인력 부족 · 분석 기술 부족
응용프로그램 보안	시큐어 코딩[2]	· 입력데이터 검증 및 보안기능 설계 · 데이터 입력, 수정, 삭제 에러처리 · DB 데이터 관리 기능

데이터 품질관리를 통한 오류개선방안과 데이터 관련 보안 정책별 보안 기술을 비교해보면 아래 [그림2]와 같이 정리할 수 있다.



〔그림 2〕 오류 개선방향 및 정책별 보안기술

1) DRS(Disaster Recovery System) : 재해복구시스템로 천재지변이나 테러 같은 참사에도 데이터를 보존하고 자동 복구하는 장치

2) 시큐어코딩 : 시큐어 코딩(SW개발보안)은 SW개발과정에서 개발자의 실수, 논리적 오류 등으로 인해 발생될 수 있는 보안 취약점, 보안약점들을 최소화하여 사이버 보안 위협에 대응할 수 있는 안전한 SW를 개발하기 위한 일련의 보안 활동을 의미[17].

데이터 저품질에 따른 품질관리 개선 활동과 데이터 보안 정책별 보안기술의 영역이 상당부분 중복되는 것을 알수 있다. 데이터 품질 환경 내 오류데이터를 보안기술을 통해 대응하여 관리한다면 공공데이터 품질을 유지해야하는 기관입장에서는 매우 효과적인 방안이 될 것이다. 이에 오류데이터의 주요발생원인에 대해 데이터보안 정책별 보안기술이 데이터 품질관리에 보완이 가능한지 여부를 알아보도록하겠다.

데이터 품질영역이 상당히 방대하고 오류데이터 발생원인이 다양함을 감안하여, 이번 연구에서는 오류유형 중 가장 비중이 높은 데이터 입력 분야의 데이터구조, 데이터값 유형에 집중하도록 하였다. 이를 통해 데이터 보안기술 중 응용프로그램 보안을 통해 데이터 입력 분야의 데이터 저품질을 향상시키고 보안위협을 제거하는 데이터 보안기술을 통한 품질관리 오류 개선 연계 대응방안을 제안하도록 하겠다.

제안에 앞서 주요 오류 발생원인별 보안기능항목을 기준으로 시큐어 코딩의 점검항목을 살펴보도록 하겠다. 시큐어코딩의 특징으로 보안기준들은 분석·설계단계에서부터 고려되어야 하며, 일부 항목들은 개발단계에 코딩 규칙을 준수하는 것만으로도 데이터 오류를 포함한 보안 취약점들을 제거할 수 있다[17].

위에서 분석했던 오류율별 주요 오류 데이터값에 대한 시큐어코딩 보안기능은 아래 [표 12]와 같다.

주요 오류데이터를 살펴보면, 여부 도메인의 오류율 총 69.95%로 가장 높았던 적재규칙 등 Null 오류의 경우 시큐어 코딩 기능 보완을 통해 이에 대한 보안으로 Null이 될수 있는 참조값을 참조하기 전에 Null 검사를 진행함으로 Null 오류 발생시키는 취약점을 제거한다.

율도메인과 코드도메인에서 나왔던 18.19%의 사전정의된 형식에 대한 오류의 경우 정수형 변수의 오버플로우는 정수값이 증가하면서, 허용된 가장 큰 값보다 더 커져서 실제 저장되는 값이 의도하지 않게 아주 작은 수이거나 음수가 되어 발생한다[17]. 언어/플랫폼별 정

〔표 12〕 주요 오류 데이터값에 대한 시큐어코딩 보안기능

주요 오류	오류율	시큐어 코딩 보안기능
Null 오류	69.95%	Null Pointer 역참조
형식오류	18.19%	정수형 오버플로우
Rule 위반	13.42%	DBMS 조회 및 결과 검증

수탑의 범위와 결과값의 범위를 체크하는 모듈을 사용하여 적절한 범위 내 값인지를 확인하는 방법으로 오류 발생시키는 취약점을 제거한다.

데이터규칙 도메인의 13.42% 오류 경우 참조값 등에 대한 사용자 규칙에 관한 오류로 DBMS 조회 및 결과 검증을 통해 응용프로그램의 외부입력값을 이용해 동적으로 처리되는 입력값에 대한 검증을 수행한 후 사용하는 방법으로 오류 발생시키는 취약점을 제거한다.

이와 같이 데이터 오류에 대한 시큐어코딩 개발보안 적용 효과를 열거하면 아래와 같다[17].

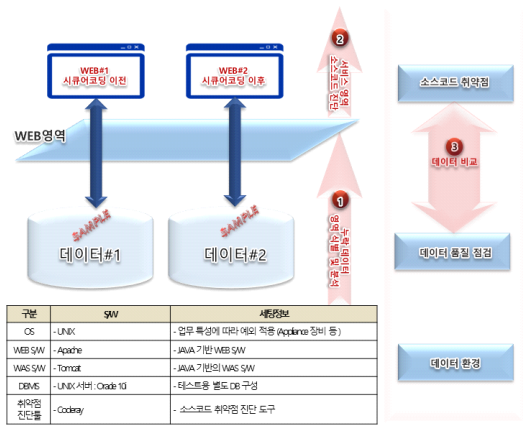
- 사용자, 프로그램 입력 데이터에 대한 유효성 검증체계 구축
- 응용프로그램에 대한 주요 발생원인을 사전에 차단하여 개선
- 연계 시스템에 대한 인증, 접근통제, 권한관리 등 데이터 품질 보안항목을 반영
- 에러 또는 오류상황을 미처리 및 추적성이 가능하도록 안전한 방안 설계

이를 통해 오류유형 중 가장 비중이 높은 데이터 입력 분야의 데이터구조, 데이터값 유형에 집중하기 위해 응용프로그램 보안에 해당되는 시큐어 코딩을 통해 입력데이터 검증 및 보안기능 설계, 데이터 입력, 수정, 삭제 에러처리를 적용할 수 있는 보안기술을 통한 대응방안에 대해서 제안하고 그 효과성을 검증하도록 한다.

4.2. 응용프로그램 보안기술 대응방안 검증

시큐어코딩을 통한 오류데이터에 대한 대응효과 검증은 기존 누적된 데이터를 기준으로 시큐어코딩이 적용된 시스템과 미적용된 시스템간의 데이터 오류 비교를 통한 방법을 통해 진행하였다. 테스트 환경은 [그림 3]와 같이 전자정부 프레임워크를 기반으로 하여 별도로 구성하였다. 테스트를 위한 데이터셋은 시큐어코딩 개선작업 이전과 이후로 나누어 두 개의 세트로 구성하였다.

테스트 대상 시스템은 2014년 시큐어코딩 개선 용역을 통해 행정안전부『소프트웨어 보안가이드 라인』을 준수한 코딩 이후 취약점 보안을 지속적으로 보완해온 응용프로그램으로 선정했다.



〔그림 3〕 테스트 방안

취약점 진단은 Trinity사의 CodeRay 소스코드 취약점 진단 시스템을 활용하여 테스트를 진행했다.

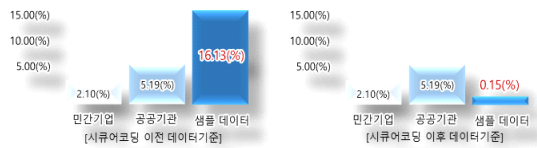
누락데이터 영역식별 및 분석에 대한 종합 결과는 아래 [그림 4]과 같이 시큐어코딩 이전 데이터기준은 16.13%, 시큐어코딩 이후 데이터 기준은 0.15%로 전체 데이터 기준 오류율은 시큐어코딩 데이터기준 오류율이 현저히 낮은 수치였으며, 이 수치는 공공기관·민간기업 평균 오류율보다 낮았다.

추가적으로 기술대응방안 검증 진단 결과에 따른 도메인별 오류건수와 오류율을 비교해보면 [그림 5]와 같다.

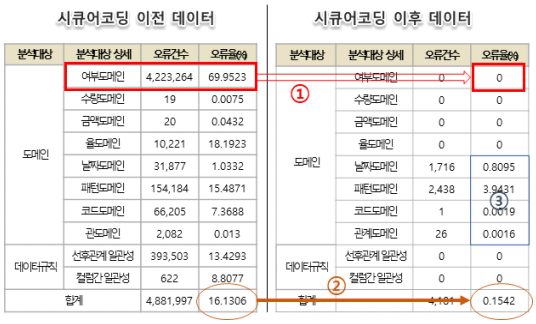
① 전체 데이터의 가장 높은 오류율을 가졌던 여부도메인에 대해서는 시큐어코딩을 통해 데이터 품질의 오류율이 0%로 보완이 된 것을 알 수 있다.

② 시큐어코딩 데이터 기준으로 테스트값을 진단한 결과는 전체 데이터 기준 오류율이 16.13%가 나온 데이터를 기준으로 테스트 데이터는 오류율이 0.15%로 현저히 감소한 것을 알 수 있다.

③ 날짜, 코드, 관계, 패턴 부분 오류율이 일부 있었으며, 패턴 적재규칙의 오류율은 3.94% 수준으로 사전 정의된 데이터 패턴 적재 규칙이 일부 누락



〔그림 4〕 기술대응방안 검증 결과



〔그림 5〕 기술대응방안 검증 진단 결과 도메인별 비교

되어 있어 사용자 오입력에 의해 오류 데이터가 쌓이고 있었으며 패턴 컬럼에 위배가 되는 데이터의 경우 검색 조건이나 통계 시에 누락될 가능성이 존재했다.해당 도메인의 경우 사용자 정의에 따라 변경이 있을 수 있는 도메인들의 값이라는 공통점이 있었다.

시큐어코딩 개선 이전 이후 데이터를 비교해보는 간단한 테스트를 통해 오류데이터의 발생원인 중 데이터 입력에 관한 오류율에 영향을 주고, 특히 파라미터에 대한 취약점 보완을 통해 기술적 대응방안의 효과가 있는 것을 알 수 있다.

시큐어코딩 보안기능항목 별 효과에 대한 세부검증이 필요하겠지만 공공데이터 품질환경 내 데이터 오류에 대한 발생원인 중 데이터 입력에 대한 기술적 대응방안에 대해 효과가 있다는 것을 알 수 있었다. 기술적 대응방안을 위해 다른 추가적인 보안기술 테스트 방법뿐만 아니라 정책적인 품질관리체계에서 관리 통제에서 보안 위협을 통해 보완할 수 있는 방안에 대해 추가적인 연구가 필요할 것으로 보이며, 추후 데이터 거버넌스, 시큐어 코딩과 데이터 품질과의 영향도 분석을 통해 데이터 품질과 정보보안 영향도에 대한 정량화를 제시할 수 있을 것이라 예상된다.

Ⅴ. 결 론

5.1. 결론을 통한 업무활용 가능성

공공데이터를 보유한 기관측면에서 저품질 공공데이터의 오류 원인에 대해 파악해서 위험 분류를 통해 효과적으로 대응하기 위한 체계적인 방안이 가장 중요하다. 본 논문에서는 단계를 나누어 오류 데이터의 원인분석을 통해 위협요인을 도출하여 체계적인 위험 분류를 통한 보안의 기술적 대응방안을 제안하였다.

첫 번째, 공공데이터를 데이터 품질 점검하여 도메인별 오류데이터를 알아보고, 오류데이터 발생원인 분석을 통해 보안위협과 공공데이터를 사용하는 사용자 측면과 기관 측면의 보안 문제를 분류하였다.

두 번째, 분류된 오류 발생원인별 보안문제를 기준으로 데이터 품질관리를 통한 개선방향을 제시하고, 품질관리 오류 개선방향별 데이터보안 정책별 보안기술을 비교하여 정리하였다.

세 번째, 보안위협을 제거하는 데이터 보안기술을 통한 품질관리 오류 개선 연계 대응방안을 제안하였다.

마지막으로, 기존 누적된 데이터를 기준으로 시큐어코딩이 적용된 시스템과 미적용된 시스템간의 데이터 오류 변화를 통해 대응방안에 대한 효과성을 검증하였다.

추후 확대될 공공기관의 공공데이터 품질 관리를 데이터보안 정책별 보안기술을 이용하면 보다 효율적인 관리를 위한 기반이 될 것이며, 앞으로 공공데이터 오류에 대한 패턴에 따른 발생원인이 다양하게 나올 것이 예상되므로 이 패턴의 다양한 분석을 통해 기관이 가질 수 있는 보안위협에 대응을 위한 연구는 실효를 거둘 것으로 기대된다. 이번 연구에서는 공공데이터의 활용을 위한 대표적인 방안인 서비스 제공을 위한 웹 서비스에 대해 오류에 대한 발생원인을 분석했지만, 그 외 다른 활용 방안에 대하여 추가적인 분석이 가능할 것으로 생각된다. 특히, 축적된 품질관리 데이터를 기준으로 통계 마이닝 기법 혹은 기계학습 기법을 통해 데이터 품질 자동판별 분석하여 오류값 범위 내의 이상값 탐지에 따른 효과적인 데이터 보안 관리체계 검증이 가능할 것으로 기대된다.

5.2. 응용프로그램 보안기술 대응방안 검증

공공데이터에 대한 오류 발생원인별 대응방안에 대해 관련분야 전문가들과의 위험분류 및 위협에 대한 관련 문헌들에 대한 고찰에 기반하였다. 그럼에도 불구하고 오류데이터에 대한 보안위협이나 보안기술에 대한 연구는 부족하여 오류에 대한 실데이터 분석을 통하여 객관성을 유지하려고 하였으나, 충분한 테스트 데이터에 대한 비판의 여지가 충분하다.

하지만 2020년부터 공공기관 데이터 품질에 대한 점검이 확대되고 점검대상 데이터가 늘어남에 따라 오류 발생원인에 대한 데이터 분석 사례도 늘어날 것이다. 또한, 데이터 품질관리를 위한 절차와 인력이 더 필요함에 따라 정보보안 기술을 이용한 관리 효율에 대한 요구도 늘어날 것으로 전망된다. 이에 따라 후속연구에서 객관성을 유지할 수 있는 충분한 사례로 보완해야 할 사항이다.

또한, 데이터 오류 발생원인의 보안위협에 대해서는 오류데이터가 다양하게 늘어남에 따라 위험의 분류 기준도 다양해지고 이를 해결하는 대응방안도 늘어날 것이라 예상된다. 이에 데이터 오류 형태와 발생원인에 대한 꾸준한 탐색과 오류별 위험 심각성 분석을 통한 위험원인별 분류, 대응방안 등 후속연구에서 다양한 분류 노력이 시도될 필요가 있다.

향후 연구를 통해 다양한 공공데이터 환경에서의 품질에 따른 연구 데이터가 충분히 누적되고, 공공데이터에 대한 품질체계과 정보보안 관리체계의 상관관계에 대한 연구가 지속된다면 보다 효율적인 정보보안체계와 데이터 품질관리체계의 통합 정보체계 마련이 가능할 것을 기대해 본다.

참 고 문 헌

〔1] “공공기관 보유 데이터 품질관리를 위한 전문 ISP”, 한국정보화진흥원, 2011

[2] “데이터 품질 가이드라인”, 한국데이터베이스진흥원, 2008

[3] “공공데이터 품질관리 매뉴얼 Open Government Data Quality Management Manual v2.0”, NIA 한국정보화 진흥원, 2018

[4] “데이터베이스 품질진단 절차 및 기법”, 한국데이터베이스진흥원, 2009

[5] “데이터베이스 품질 검사항목 : DQC-M”, 한국데이터베이스진흥원, 2010

[6] “데이터베이스 품질 검사항목 : DQC-V”, 한국데이터베이스진흥원, 2010

[7] “데이터베이스 보안 검사항목 : DQC-S”, 한국데이터베이스진흥원, 2010

[8] “데이터 품질지표”, Kerr, 2006;Batiniet al, 2009;Madnick, 2009

[9] 윤소영, “공공데이터 활용을 위한 링크드 데이터 국가 연계체계 구축에 관한 연구”, *정보관리학회지* 30(1), pp. 259-284. 2013

[10] 정승호, 정덕훈, “공공기관의 데이터 품질에 영향을 미치는 요인에 관한 연구“, *정보처리학회논문지* 2, pg. 251, 16p. 2013

[11] 김현철, 김광용, ”공공 데이터 품질이 공공 데이터 개방의 신뢰에 미치는 영향에 관한 연구“, *한국IT서비스학회지* pp 53-68, 2015

[12] 장경애, 김자희, 김우제, ”데이터 공학 고객의 요구사항에 기반한 데이터품질 평가속성 및 우선순위 도출”, *정보처리학회논문지* pg 549, 12p 2015

[13] 윤승영, “기업지배구조 관점에서 바라본 내부통제와 기업의 정보보안”, 기업법연구 30(1) pp 9-37, 2016

[14] 장경애, 김우제, “Data Governance 평가를 위한 속성지표 연구”, *정보처리학회논문지* pg 57, 10p, 2017

[15] “공공데이터 품질관리 수준 평가지표, 한국정보화진흥원, 2015

[16] ”공공데이터 품질관리 매뉴얼”, 한국정보화진흥원, 2018

[17] “전자정부 SW 개발·운영자를 위한 소프트웨어 개발보안 가이드”, 행정안전부, 2019

〈저 자 소 개〉

이 원 재 (LEE Won Jae)

정회원

2002년 2월 : 고려대학교 전산학과 학사

2018년 9월~현재 : 고려대학교 사이버보안학과 석사과정

<관심분야> 정보보호정책, 데이터 보안, 데이터 거버넌스

김 휘 강 (Huy Kang Kim)

종신회원

1998년 2월 : KAIST 산업경영학과 학사

2000년 2월 : KAIST 산업공학과 석사

2009년 2월 : KAIST 산업및시스템공학과 박사

2004년 5월~2010년 2월 : 엔씨소프트 정보보안실장, Technical Director

2010년 3월~현재 : 고려대학교 정보호호대학원 교수

<관심분야> 온라인게임 보안, 네트워크 보안, 네트워크 포렌직, 침입탐지시스템, 봇넷탐지