DQC 데이터베이스품질인증
Database Quality Certification

데이터 인증을 통한 데이터 품질확보 전략
(DQC-V 가이드라인 개정 발표)

지티원 DG서비스사업부
박선종
2013.10.08

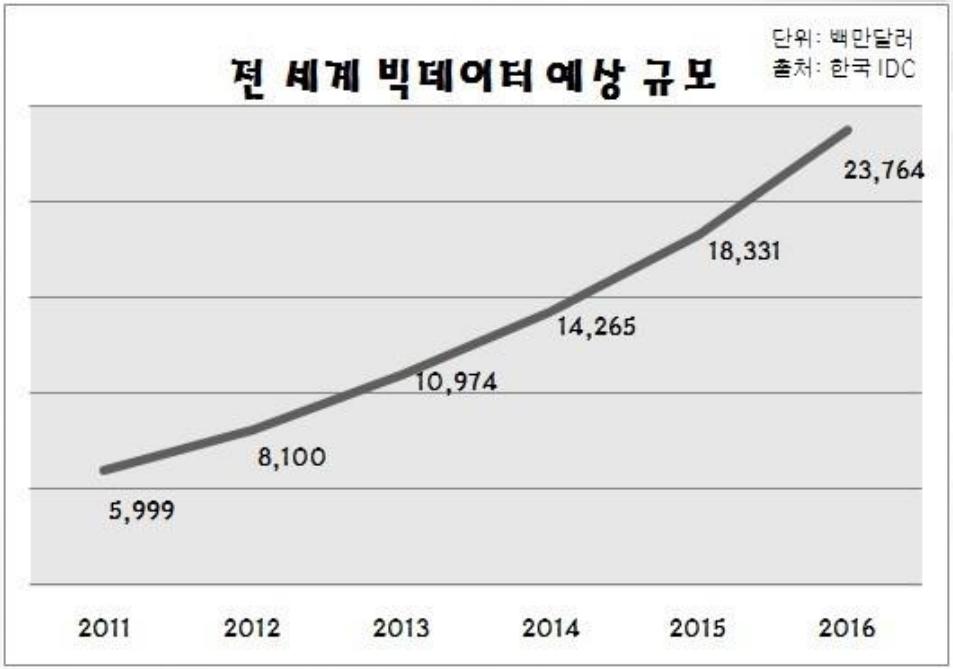
GTONE
Governance Technology

Contents

데이터 관련 주요 트랜드 및 이슈

빅데이터 환경 도래

시스템, 서비스, 조직(회사) 등에서 주어진 비용, 시간 내에 처리 가능한 데이터범위를 넘어서는 데이터



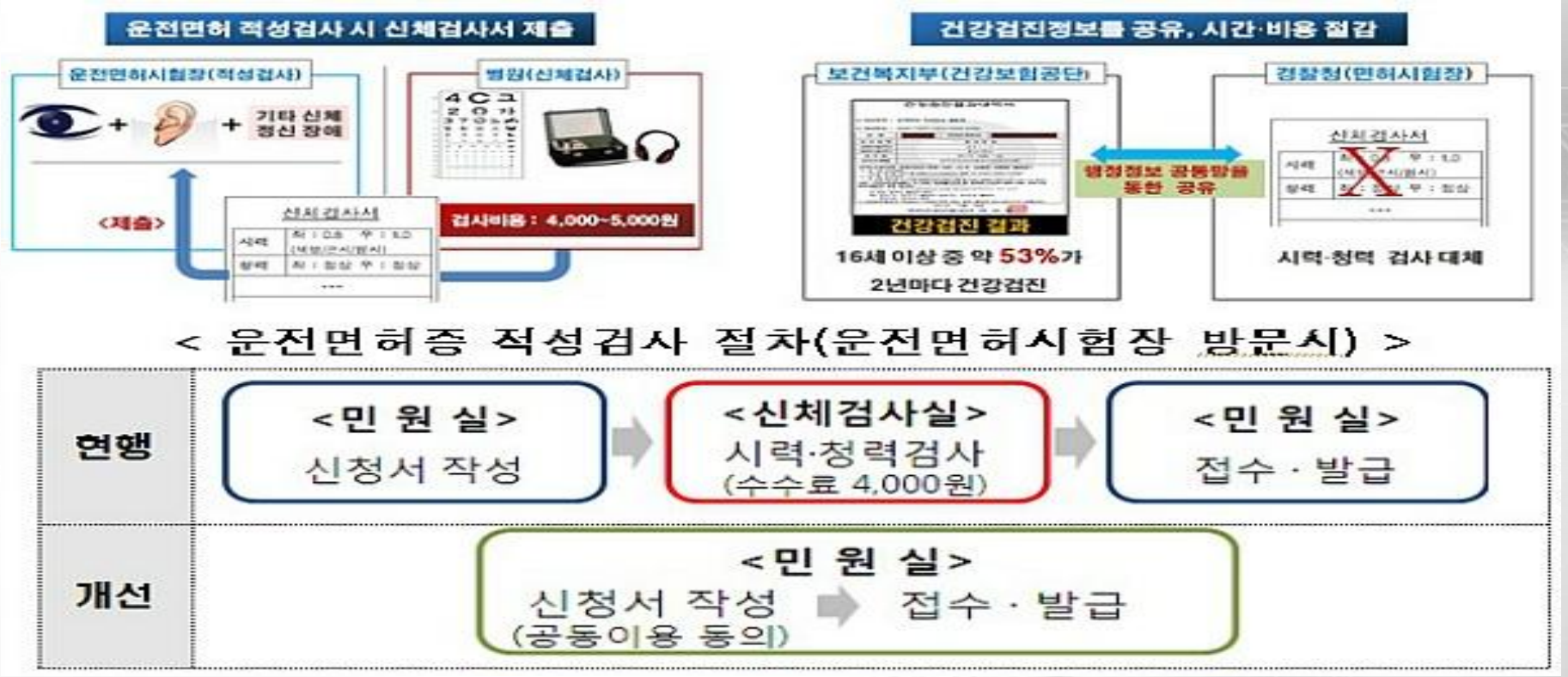
출처 : 한국IDC

출처 : IDC 2011

데이터 관련 주요 트랜드 및 이슈

공공정보 공유 및 개방 가속화

경찰청-건강보험공단 간의 정보공유로 운전면허 적성검사가 간소화 함



출처 : 행정정보 공유 활성화 로드맵

데이터 관련 주요 트랜드 및 이슈

공공부문 데이터품질관련 사업 활성화

- '13년 공공데이터 **품질관리 지원사업 추진**
  - 1차(품질진단 및 개선) : 18개 기관 / 2차(품질진단) : 17개 기관
- 국가 공공DB 품질개선 추진 체계 마련('13.하반기)
  - 중요DB개정 및 연차별 품질지원 계획 수립
  - 품질관리 현황의 주기적 조사 및 현행화 등 관리체계 마련
- 공공DB 품질관리 법제도 개선 및 관련 연구 추진
  - 공공데이터법 시행에 따른 시행령 규칙 관련 지침 제개정 지원
  - 공공데이터 거버넌스, 자가진단 및 대가산정 기준 개선 등
- 품질관리지원센터* 운영(연중 지속) * http://www.gooddata.kr
  - 기관별 맞춤형 컨설팅 및 커뮤니티 운영 등
- 교육 및 홍보 추진(연중 지속)
  - 지역 거점별 순회 교육 및 교육원 정규 과정 반영 등

데이터 관련 주요 트랜드 및 이슈

인식의 변화 및 관리 솔루션간 통합·연계 기능확대

1. 무엇을 → 어떻게

기관 담당자 데이터의 인식 변화

- 일회성, 휘발성 데이터 관리의 아닌 지속적이며 체계적인 데이터 관리를 위한 방안 모색 중
- 관련 법제도 개선 및 대내외 업무환경 변화에 대비한 데이터 관리 방안 모색

2. 기능 → 프로세스

유틸리티 혹은 솔루션에서 시스템으로 변화

- 메타데이터 및 품질관리 뿐만 아니라 데이터 흐름정보(ETL,EAI), JOB정보, 성능관리, ITSM 등과의 통합 및 연계로 발전 중
- 기능중심의 화면에서 프로세스 중심의 화면으로 전환 중
  예) 데이터관리 포탈

데이터 관련 주요 트랜드 및 이슈

기관내부의 데이터관리 주요 이슈

지속적인 정보시스템의 발전 영향

- **시스템 교체**
- 신규 개발
- **기업 환경의 변화** (제품 라인 변화, 시장 변화, 비즈니스 변화)
- 통신,스토리지,소프트웨어 기술 발전

부정확한 데이터 허용

- **데이터 베이스 중요성의 비해 부정확한 데이터에 대한 낮은 인식**
- 시스템 변경 속도가 빨라 통제 어려움
- 데이터 문제에 대한 비용 개념이 낮음

데이터 관리 기술부족

- **데이터 관리 전문가 부족**
- 데이터 품질 관리 솔루션 의 적용 사례가 부족
- **각 기업(기관)에 적합한 데이터 품질 활동을 고려한 적용 기술 부족**

R&R 기반의 조직문화

- **데이터 보정, 재작업 등을 정상적인 부분으로 인식**
- **비난을 우려해 이슈화 하지 않음**
- 데이터 품질 관리 비용의 인식 부족

데이터 관련 주요 트랜드 및 이슈

해결 방안?

대외 환경 이해

- 빅데이터 환경 도래
- 공유 및 개방 가속화
- 고품질의 데이터 요구

데이터 관리

1. 무엇을, 어떻게 할 것인지?

2. 조직내부의 협업과 단기적인 성과를 볼 수 있는 방안은 있는지?

내부 문제

- 부정확한 데이터 지속적 발생
- 데이터 관리 기술 및 인원 부족
- 성과중심의 조직 문화 (엄격한 R&R)

Chapter 2

데이터베이스 품질인증 소개

데이터 인증(DQC-V)

• 도메인, 업무규칙을 기준으로 데이터 값 자체에 대한 품질 영향요소 전반을 심사·심의하여 인증하는 것을 의미함

데이터관리 인증(DQC-M)

• 정보시스템에 대한 데이터관리 체계를 심사하여 인증하는 것을 의미함

데이터 보안 인증(DQM-S)

• 데이터베이스를 대상으로 접근제어, 암호화, 취약점분석, 작업결재 등 데이터베이스 보안에 대한 기술요소 전반을 심사하여 인증하는 것을 의미함

Chapter 2

데이터베이스 품질인증 소개

인증심사 대상

데이터 인증(DQC-V)의 심사대상은 행정 및 업무지원, 의사결정 및 정책지원, 지식공유 및 활용 등 임의의 목적을 위해 구축된 데이터베이스

인증심사 수준 평가

데이터 인증(DQC-V)은 DB내 값(VALUE)를 대상으로 데이터 정합성(반·오류율)을 정량화하여 정합성 수치에 따라 인증 수준을 결정

구분	정합성율	비고
Platinum Class	99.977% 이상	5.0시그마 이상
Gold Class	97.700% 이상	3.5시그마 이상
Silver Class	95.510% 이상	3.2시그마 이상

Chapter 2

데이터베이스 품질인증 소개

인증심사 기준

심사영역	심사항목	심사내용
도메인	번호	패턴 및 체크비트
	금액	허용범위
	명칭	패턴
	수량	허용범위 및 단위
	분류	표준정의
	날짜	허용범위 및 날짜 값
	비율	허용범위
	내용	적용언어 패턴
	코드	허용 코드 값
	키(Key)	참조무결성
	공통	표준준수 여부

심사영역	심사항목	심사내용
업무규칙	관계자	주제영역별 업무규칙 준수 여부
	상품	
	계약	
	활동	
	거래	
	자원	
	지원	
	생산	

Chapter 2

데이터베이스 품질인증 소개

1. 도메인 기반 데이터품질진단

Profiling

컬럼 분석 (Column Analysis)	• 통계적 기법을 통한 데이터 분석 : 표본추출~전수검사 → Null 개수, Space 개수, Min, Max, 평균, 표준편차,분산 등
날짜 분석 (Date Type Analysis)	• Data Type은 Character이나 의미상 날짜/시간 유형 데이터에 대한 유효성 분석 → 고객의 생월일에 대한 MMDD 날짜유형 오류 분석
패턴 분석 (Pattern Analysis)	• 문자, 숫자 등으로 구성된 특정 패턴을 갖는 값이 일관된 패턴을 갖는지 여부 점검 → 주민등록번호 숫자13자리
코드 분석 (Code Analysis)	• 개별코드/통합코드 마스터와 트랜잭션 코드와의 유효성 분석 →고객의 직업세부코드
참조무결성 분석 (Referential Integrity Analysis)	• 부모-자식 관계 데이터의 참조 무결성 분석 → 고객의 고객번호

Domain Group

- Indicator
- 금액
- 단위
- 리스트
- 포맷
- 코드
- 날짜
- 명칭
- 건수
- 수량
- .....

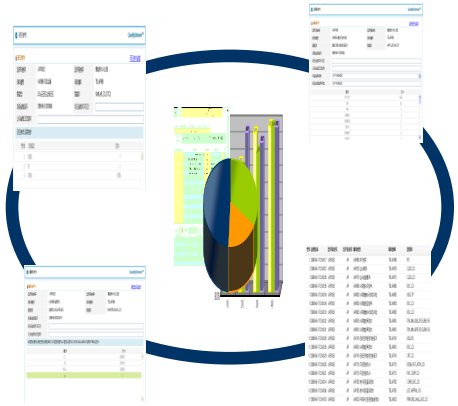
Chapter 2

데이터베이스 품질인증 소개

2. 업무규칙 기반 데이터품질진단

기술적 접근에 의한 업무규칙 설계(Inside - Out)

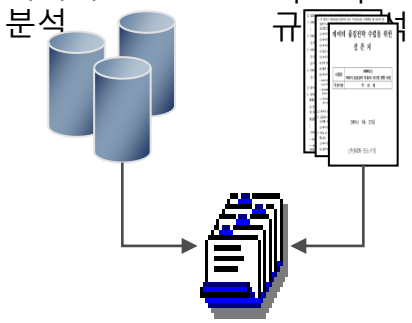
프로파일링 보고서 및 분석 자료



1) 컬럼 분석
2) 비정형패턴 분석
3) 날짜유형 분석
4) 코드 분석 및 참조무결성 분석

업무적 요구사항에 의한 업무규칙 설계(Outside - In)

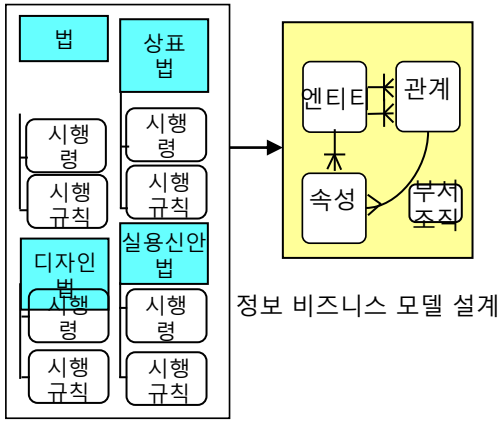
업무적 요구사항 분석



1) 검증목표데이터 맵 분석
2) APP. 비즈니스규칙 분석
3) 표준업무규칙_기술적 DR 설계

정보 비즈니스 모델에 의한 업무규칙 설계(순수 근거규정에 근거)

법령, 규정,업무지침 분석



1) 법령, 규정, 지침분석
   - 근거규정 : 법, 시행령, 시행규칙 등을 근거
   - 정보비즈니스 모델 설계

Chapter 2

데이터베이스 품질인증 소개

3. 진단결과

- **완전성** — 단독완전성, 조건완전성
- **일관성** — Table간 일관성, 컬럼간 일관성
- **유효성** — Range 유효성, Data time 유효성, 포멧유효성, 코드유효성
- **유일성** — 유일성 Ex) 시스템ID 사용시 PK는 유일해야 한다.

대상		품질평가지표				품질측정결과(오류 발생분)						DPMO	시그마 수준	품질지수
		완전성	유효성	일관성	유일성	총건수	완전성	유효성	일관성	유일성	소계			
Tables	컬럼	Y	Y	N	Y	1,000	0	0	0	0	0	0	6.9	100
Table	컬럼	Y	Y	N	N	44,782	44,782	44,782	0	0	89,564	2,000,000	0	0
Table	컬럼	Y	Y	Y	Y	10,000	0	200	0	0	200	20,000	3.6	73
합계														

- 측정 대상 테이블/ 컬럼별로 정의된 DQI 에 따라서 측정을 수행하여 위의 결과보고서를 Reporting 한다.

Ex) 유효성 오류유형 - 여신원장

대출계좌번호	발행일	만기일	상태	유형
111-111-1111	2006-01-31	2005-12-31	정상	발행일>만기일
222-222-2222	0001-01-01		해지	만기일 불완전
333-333-3333	2006-03-11	9998-12-01	정상	Out of Range
444-444-44444	20060131	20071231	정상	Format 오류

Chapter 2

데이터베이스 품질인증 소개

주요 준비사항

단계	준비사항	목적
인증상담	• 정보시스템 및 데이터베이스 개요	• 심사대상 데이터베이스에 대한 현황 파악 및 심사 범위, 심사기간 수립
인증신청	• 테이블·컬럼 정의서, ERD, 코드정의서, 데이터 입력·가공 지침 등	• 심사대상 데이터베이스의 도메인 분석
인증심사	• 데이터 관리 정책, 데이터 입력 및 가공지침, 데이터 표준 정의서 관련 문서 및 운영 현황	• 심사대상 데이터베이스의 품질 확보를 위한 관리절차 수립 여부 및 운영 현황 분석
인증유지	• 정보시스템 운영 현황 및 최근 인증의 인증심사보고서	• 재인증 신청대상의 동일성 여부 검토 및 심사기간 수립

Chapter 2

데이터베이스 품질인증 소개

데이터베이스품질인증 사이트
http://www.dqc.or.kr

인증절차

Chapter 2

데이터베이스 품질인증 소개

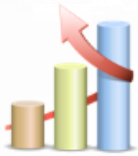
DQC-V 인증을 통한 기대효과

데이터관련
노하우 전수

+

공감대 형성 및
동기 부여

대외적인 공신력
확보



데이터 가치
창조

Chapter 3

DQC-V 데이터품질가이드라인 주요 개정 내용

주요 요청사항

- 도메인 기반
  - 분석 스크립트 파일요청
  - 분석대상 DBMS 추가 요청
  - 분석대상 ERD 파일 요청
- 업무규칙 기반
  - 업무규칙 도출 세부 방안 요청
  - 주제영역별 전체 ERD 파일 요청
- 공통 사항
  - 오타 및 분석 SQL 수정 보완

개정내용 범위 및 방향성

- 진단스크립트 파일 제공
  - ORACLE 및 MS-SQL DBMS 에서 수행 가능한 SCRIPT 제공
- 업무규칙 도출 방법 추가
  - 기술적, 업무적, 법령 기반 등
- 자가진단 할 수 있는 정도의 스크립트 상세화
  - 테이블생성, 데이터 입력 등

Chapter 3

DQC-V 데이터품질가이드라인 주요 개정 내용

구분	개정 및 보강 내용
1장 데이터 품질 이해	• BR도출 방법에 대한 내용 보완 - 법령, 업무규정 분석에 따른 업무규칙 도출과정 및 기술적 도출과정 내용 추가 • 품질통제 프로세스 추가 • 품질진단 스크립트 보강 - ORACLE 및 MSSQL 수행 할 수 있는 SQL추가 - 정규표현식
2장 1절 도메인	• 날짜 유형 분석 시 SQL사용 형식을 기준날짜 매칭방식과 유효성 검증 로직 방식을 둘다 표기하고, 패턴분석은 패턴분석 함수를 제시하여 활용 • 1장에서 제시한 함수를 활용해서 도메인 진단 내용 보완 및 상호 내용 일관화 • 최신 스크립트를 사용해 SQL 구문 보강
2장 2절 업무규칙	• 전체 주제영역을 포함하는 ERD작성(4절지)및 세부 주제영역별 모델 기반 업무규칙 도출 사례 • 주제영역별 업무규정 및 시행령 기준 업무규칙 도출 방안

Chapter 4

발표 요약

- 데이터관리의 인식을 정성적, 정량적으로 인식 전환
- 데이터관리는 산이 아니라 지속관리 관리와 활동이 필요한 정원

DQC 데이터베이스품질인증
Database Quality Certification

Thank You!

GTONE
Governance Technology