Journal of the Korean Society of Marine Environment & Safety
Technical Report
Vol. 29, Special Issue, pp. 067-074, October 31, 2023, ISSN 1229-3431(Print) / ISSN 2287-3341(Online) https://doi.org/10.7837/kosomes.2023.29.s.067

위험유해물질 데이터 품질 향상을 위한 표준화 연구

오진덕* · 김주영**† · 김민섭*** · 김용명**** · 이문진*****

* ㈜온더시스 사장, ** ㈜온더시스 시스템 개발부 이사, *** ㈜온더시스 기술영업부 부장,
**** 한국해양과학기술원 부설 선박해양플랜트연구소 시험기술원,
***** 한국해양과학기술원 부설 선박해양플랜트연구소 영년직 책임연구원

Standardization Study to Improve HNS Data Quality

Jinduck Oh* · Juyeong Kim**† · Minseob Kim*** · Moonjin Lee**** · Yongmyung Kim*****

* President, Onthesys Inc., Daejeon, 34025, Korea
** Associate Executive Director, System Development Department, Onthesys Inc., Daejeon, 34025, Korea
*** Head of Department, Technical Sales Departmentm Onthesys Inc., Daejeon, 34025, Korea
**** Researcher, Korea Research Institute of Ships & Ocean Engineering, Daejeon, Korea
***** Senior Researcher, Korea Research Institute of Ships & Ocean Engineering, Daejeon, Korea

요 약 : 우리나라는 삼면이 바다로 이루어져 있고, 이에 따라 많은 해양 시설로 인한 위험유해물질이 배출되고 있으나, 배출관리 및 규제 시스템이 미비한 상황이다. 따라서, 위험유해물질(HNS) 관리를 위하여 효율적으로 데이터를 수집할 수 있는 시스템이 필요하다. 본 연구에서는 HNS 데이터를 효율적으로 관리 및 저장하기 위한 데이터 표준화 시스템을 설계하고 이의 표준화 방안을 제시하고자 한다.

핵심용어 : 위험유해물질, 데이터베이스, 데이터 표준화, 데이터 품질, 표준 도메인, 표준 단어, 표준 용어

Abstract : *Korea, is bounded on three sides by the sea, into which dangerous substances are discharged from many marine facilities.; However, the emission management and regulatory systems are insufficient. Therefore, there is a need for a system that can efficiently collect data for HNS management. In this study, we designed a data standardization system for efficiently managing and storing HNS data and proposed a standardization plan.*

Key Words : *HNS, DB (Database), Data Standardization, Data Verification, Standard Domain, Standard Word, Standard Terminology*

1. 서 론

1.1 연구배경

풍부한 해양 생태계를 가지고 있는 청정 바다를 보유한 우리나라는 3면이 바다에 둘러싸여 있는 나라이며 그에 따라 다양한 종류의 많은 해양 산업시설을 보유하고있다. 이 다양한 많은 해양 산업 시설에는 해양 유입 가능성이 높은 위험·유해물질(HNS)을 가지고 있지만 HNS(Hazardous & Noxious Substance) 배출 관리를 전면 해양 배출금지하고 있고 이로 인해 사회적인 이슈 및 갈등이 발생하고 있고 이에 따라 산업계와 환경계의 요구를 합리적으로 충족시킬 수 있는 합리적인 HNS 배출 관리 규제 마련이 필요하게 되었다.

각 해양 산업 시설이나 해안가에 합리적인 HNS 배출 관리 규제 마련을 위해 의사결정의 핵심요소로 대두되는 HNS 데이터를 데이터 통합 및 데이터 품질에 대한 관심이 증대됨에 따라 HNS 데이터 품질 향상을 돕기위한 HNS 데이터 표준화 시스템 모델을 제안하였다.

오늘날 정보 제공 기반의 다각화가 심화됨에 따라 데이터의 정보화 및 고도화 요구가 지속적으로 증가되고, 데이터의 깊이와 양은 급속하게 증가되고 있다. 또한, 정보의 가치에 대한 인식 변화에 따라 이용자들은 정보서비스의 가치, 품질 등에 민감해지고 있다(Korea Database Promotion Center, 2017).

* First Author : ojduck@onthesys.com, 042-484-2013
† Corresponding Author : kjy0923@oonthesys.com, 042-484-2013

이에 따라 대용량 데이터의 품질은 점차 데이터 경쟁력 확보를 위한 이슈로 등장하고 있다. 그러함에도 불구하고 대부분의 데이터는 Poor Data Quality로 인해 많은 어려움을 갖는다. 데이터 품질이 데이터의 경쟁력에 영향을 주는 핵심 요소임에도 불구하고, 현 정보시스템 현실에서는 데이터 품질 저하라는 심각한 상황을 맞고 있다. 데이터의 품질 저하로 인한 비용이 전체 제조업의 내부 실패 비용(손실 및 재작업 비용)의 40~60%를 차지하고 있다는 연구 결과까지 보고되고 있다(English, 1999).

이제 데이터의 품질은 데이터의 전 측면에 걸쳐 중요한 영향 요인으로 작용하고 있으며, 정보화 시대의 데이터에 있어 중요한 경쟁력의 척도로써 작용하고 있다. 이에 따라 데이터 품질 관리에 대한 연구들이 활발하게 이루어지고 있으며, 실제 실무에서도 관련 Tool들이 일부 활용되고 있는 추세이다.

이처럼 데이터의 품질의 중요성을 고려할 때, 데이터 구조 설계가 가지는 책임을 간과하지 않아야 한다. 근본적으로 잘못된 데이터 구조 설계에 의해 만들어진 데이터는 비록 그 데이터 값 자체는 실물과 일치되어 입력된다 하더라도 품질이 좋은 데이터라고 할 수 없다. 데이터 구조의 근본적인 문제점을 가진 데이터들은 지속적으로 품질이 낮은 데이터를 생산해 낼 수 밖에 없다는 한계점을 갖게 되기 때문이다(Kim and Park, 2004).

이에 따라 데이터의 구조적 품질 수준을 높이기 위한 연구가 필요하다. 본 연구에서는 데이터 표준화 개념을 적용하여 데이터의 구조적 품질 수준을 높일 수 있는 방안을 제시하였다. 제시된 방안은 데이터베이스 설계와 병행하여 표준 데이터를 구축하고 이를 통해 자연스럽게 데이터 표준화 작업이 수행될 수 있도록 함으로써, 데이터베이스 품질 뿐만 아니라 설계의 생산성을 향상시킬 수 있는 실무적인 적용 기반을 마련하였다.

2. 관련 연구

2.1 데이터 품질 개념과 특성

데이터 품질(Data Quality)의 정의를 Larry P. English는 데이터와 고객의 목표를 달성하기 위해 데이터 대한 이해 관계자들의 기대감를 만족시키는 것이라고 정의하였다. 또한, 과학적 기법을 통해 데이터 품질 유지를 위한 원칙으로 고객에 집중하여 데이터에 대한 성능향상 활동을 수행하도록 제시하고 있다(English, 1999).

소프트웨어 품질 특성과는 달리 데이터 품질 특성(Data Quality Characteristics)은 표준이 명확히 정립되어 왔다. 그 중 대표적인 것으로 Wang의 연구를 들 수 있는데, 데이터품질은 6가지 차원, 즉 신뢰성(Reliability), 기능성(Functionality), 효율성(Efficiency), 이식성(Portability), 사용성(Usability), 유지보수성(Maintaina bility) 등으로 구분한다(IOS/IEC Standard, 2000).

각 데이터들의 품질에 대한 측정은 소프트웨어 품질 측정 표준 도구인 ISO/IEC 9126와 ISO/IEC 14598을 기반으로 ISO/IEC 25000 데이터들 품질 특성을 측정하기 위한 방안들에 대한 연구들이 진행되고 있다.

데이터 오류를 분류 및 분석하고, 오류들로부터 데이터들 품질 특성을 측정하기 위한 메트릭들을 제시하였다.

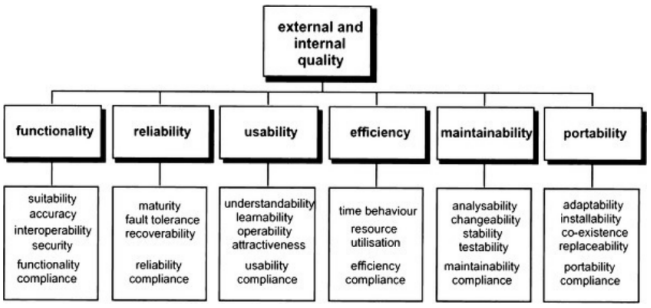

Fig. 1. Software quality requirements.

2.2 HNS 데이터의 표준화

데이터 표준화는 단위 시스템별로 산재되어 있는 도메인에 대한 표준 원칙 및 데이터에 대하여 명칭을 수립하여 표준 데이터를 구축한 후 전체적인 시스템 적용하는 방법으로, 데이터의 품질 특성을 높일 수 있는 현실적인 방안이다. 표준 데이터란 정보 시스템에서 사용하는 도메인, 코드, 용어 및 기타 데이터 관련 요소들(Elements)에 대해 공통된 내용과 형식들로 정의하여 사용하는 표준 관련 데이터를 의미한다(Korea Database Promotion Center, 2017).

미 국방부에서는 통합되고 효율적인 방식으로 미션을 수행하기 위한 DoD 8320.1-M-1 'Data Element Standardization Procedure'를 발표하였다. DoD 8320.1에서는 데이터 관리 정책 및 표준 데이터를 승인, 개발, 구현, 유지보수 하는 절차를 제시하였다(DoD4000, 1996).

데이터베이스 통합을 위한 데이터 표준화 방안을 제안하였다. 이 연구에서는 데이터 표준화 절차 및 지침과 데이터 표준화의 필요성이 제시되어 있다.

데이터의 구조적 품질 관리 성숙도 모델을 제시하였다. 이 모델은 Level 1에서 Level 4까지의 4단계로 구성되어 있으며, 상위 단계로 올라갈수록 데이터의 구조적 품질 관리 수준이 성숙된다고 정의하였다(Kim and Park, 2004)

이 연구에서는 표준 데이터를 선 구축한 후 신규 시스템 개발 시 참조하도록 한 표준 데이터베이스 관리 시스템 구축 사례가 제시되어 있다.

2.3 HNS 데이터 품질관리

데이터 품질 관리란 조직이나 기관 내외부의 정보 시스템 및 데이터베이스 사용자들의 기대감을 만족시키기 위해 지속적으로 수행하는 데이터 개선 및 관리 활동을 의미한다. 데이터 품질 관리에 대한 연구 초기에는 품질 측정에 대한 현상 분석 중심이었으나, 점차적으로 품질 성능 향상을 위한 모델 중심으로 바뀌고 있다.

MIT의 TDQM(Total Data Quality Management) 프로그램은 데이터품질을 표현적(Representational), 본질적(Intrinsic), 연관적(Contextual) 품질 등의 범주로 분류하고 각 범주 별로 데이터 품질 이슈가 발생하는 개선방안 및 패턴(Pattern)을 연구하고 있다(Kulpa and Johnson, 2003).

Larry P. English의 TIQM(Total Information Quality Management) 모델은 프로세스를 6단계로 구성하고 있으며 각 단계의 프로세스들은 유지보수를 위한 프로세스 개선 및 조직을 데이터 품질의 문화로 전환 하는 프로세스, 데이터 이행 통제, 평가, 유지보수 프로세스 로 구성된다.

국내 데이터베이스 진흥센터에서는 데이터 품질관리지침을 통해 품질 개선을 위한 프레임워크를 제시하였다. 데이터 품질 관리 프레임워크는 정보시스템의 데이터 품질 확보를 위한 필수 요소로 Enterprise Architecture 의 개념을 도입한 것으로 표준 데이터를 정의하였다.

3. 데이터 표준화 구현 방안

3.1 데이터 표준화 연구 방향 및 범위

데이터 품질에 대한 2장에서의 선행 연구들을 종합해 보면, 데이터 품질을 향상시키기 위한 여러 가지 개선 방안들이 제시되고 있음을 알 수 있다. 최근의 연구들은 관리 프로세스나 평가 측면에 중점을 두는데, 이는 제품의 품질을 향상시키기 위해서는 프로세스 품질 향상이 선행되어야 한다는 소프트웨어 프로세스 평가 모델인 CMMI(Ca pability Maturity Model Integration)의 사상과도 부합된다(Kulpa and Johnson, 2003)

그러나, 기존의 데이터 품질 개선에 대한 연구는 현상적인 데이터의 품질 개선에 한정되거나, 개선 방안의 구체성(具體性)이 부족한 측면이 많아 실무적 적용에 한계를 가지고 있다. 이에 따라 실제 실무에 적용하여 데이터 품질 관리 수준을 높이기 위해서는 보다 구체적인 방안이 제시될 필요가 있다. 또한 기존 연구의 데이터 표준화인 AS-IS 데이터를 이용하여 전사적인 Top-Down 방식으로 표준 데이터를 구축한 후 각 단위 시스템에 적용하는 방식이어서 데이터베이스 통합 또는 프로젝트 초반부터 표준 데이터 구축을 위한 자원을 확보하기가 어렵거나, 차세대 프로젝트 등에 적용하기에는 유리한 AS-IS 데이터가 없는 보통의 실무에 적용하기에는 많은 어려움을 갖게 된다.

이에 따라 본 연구는 기존 연구의 데이터 표준화 개념을 적용하되 절차를 개선하여 표준 데이터 구축을 데이터베이스 설계와 병행하여 수행 할 수 있는 체계적인 구현 방안과 사례를 제시하였고, 데이터베이스 품질 뿐만 아니라 설계의 생산성을 향상시킬 수 있는 실무적인 적용 기반을 마련하였다.

본 연구에서의 Table 1과 같이 표준 데이터를 기본으로 물리 컬럼과 논리 속성까지 데이터 표준화 범위에 포함시켰다(Korea Database Promotion Center, 2021).

Table 1. Data standardization component

Classification	Sortation	Explanation
Standard Data	Standard Data	the smallest unit of meaning that constitutes a standard term
	Standard Domain	Groups with consistency in data formats
	standard terminology	a combination of standard words that are independent and have specific meanings
	Standard Code	Symbols for shaping data values
Model Data	Logical Properties	Attribute of the entity
	Physical Properties	Meaning the column in the table

3.2 데이터 표준화 환경 및 관리 프로세스

(1) 데이터 표준화 환경

표준 데이터 구축을 데이터베이스 설계와 병행 하기 위해서는 표준 데이터 정보와 설계 정보가 서로 공유될 수 있도록 하여야 한다.

Fig. 2에서 제시된 것처럼 각 메타 정보에 대한 Repository를 공유할 수 있는 환경이 필요하다.

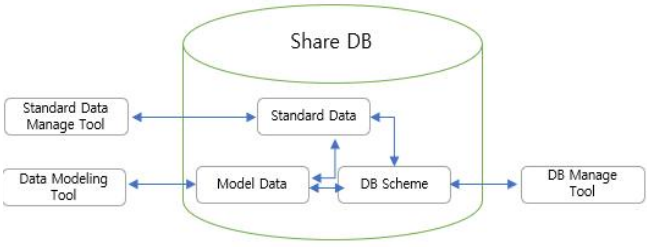

Fig. 2. Environment for Data Standardization.

(2) 데이터표준화 관리 프로세스

데이터 표준화와 데이터베이스 설계가 병행하기 위해서는 엄격한 관리 프로세스 유지가 필요하다. 또한 프로세스 유지를 위한 설계자들의 준수노력과 DA(Data Architect)의 역량이 함께 요구된다. Fig. 3에서 제시된 것처럼 논리 엔티티의 속성 설계 시작부터 표준 데이터가 발생할 수 있도록 한다.

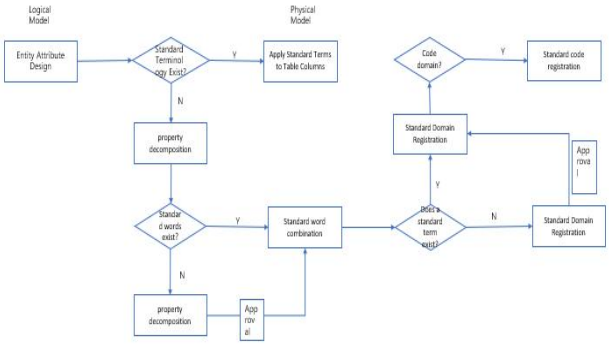

Fig. 3. Management process for data standardization.

3.3 데이터 표준화 설계지침

(1) 표준 단어 설계

① 모든 표준 단어에 단어가 의미하는 내용을 정의한다.
② 명사형으로 정의하고, 특수기호를 포함하지 않는다.
③ 업무적으로 사용되는 단어 및 표준어를 사용한다.
④ 분류 단어와 기본 단어로 구분한다.
⑤ 동의이음어는 대표 단어를 선정하여 사용한다.
⑥ 하나의 영문 약어에 하나의 한글 표준 단어를 정의한다.
⑦ 분류 단어는 데이터의 표현 방법을 나타내므로 불분명한 저장 형태는 배제한다.(배제 예: 전화, 결과,의견 등)

(2) 표준 도메인 설계

① 값, 수나 금액 성격의 도메인은 데이터 길이 유형별로 각각의 하위 도메인을 생성할 수 있다.(예 : 값, 거리, 수, 계수 등)
② 도메인명은 데이터의 속성 및 타입을 인식할 수 있는 용어로 선정한다.(예: 번호, 일자, 코드, 수, 명 등)
③ 하나의 표준 도메인에 하나의 데이터 타입 및 길이를 부여한다.
④ 코드 도메인과 번호는 표준용어별로 각각의 하위 도메인을 생성할 수 있다.(예: BOD 측정값, COD 측정값 등)

(3) 표준 용어 설계

① 하나의 도메인을 하나의 표준 용어에 결합한다.
② 두 개 이상의 표준 단어로 표준 용어는 구성된다.
③ 표준 용어의 영문명은 표준 단어에 적용된 영문 약어를‘-’를 이용하여 조합한다.
④ 하나의 표준 용어는 유일해야 한다.
⑤ 표준 용어는 Fig. 4처럼 분류어, 주제어, 수식어를 조합하여 만든다. 용어의 맨 우측은 데이터의 표현 방법을 나타내는 분류 단어(분류어)이어야 한다.
- 수식어의 예 : 배출, 유입, 발생, 보류, 승인 등
- 주제어의 예 : 수질, 오염, 농도 등
- 분류어의 예 : 비율, 빈도, 기간, 수, 값 등
⑥ 용어가 의미하는 내용을 기술하여, 동일한 의미를 갖는 용어들은 하나의 표준 용어로 정의한다.

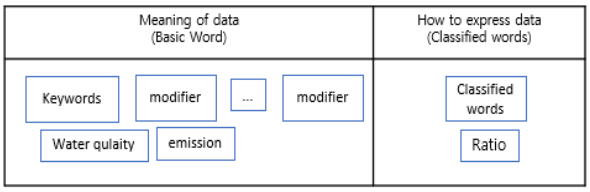

Fig. 4. Components of Standard Terminology.

(4) 표준 코드 설계

① 표준 코드의 정의는 코드 설계서에 기술한다.
② 코드 도메인이 적용된 표준용어에 대해서는 표준 코드를 정의한다.

(5) 논리 엔티티 속성 및 물리 테이블 칼럼 설계

① 적용된 표준 용어의 영문명을 물리 테이블 칼럼명으로 그대로 사용한다.
② 설계 예: ‘수질’인 속성을 설계
- 표준 단어 : 수은(MRC), 수질(WTQLT) 등
- 표준 도메인 : PH(PHDEC)
- 표준 용어 : 12월 수질 측정 값 (MON12_WTQLT_MESURE_VU)
- 테이블 칼럼 : WTQLT_PHDEC_MRC NUMBER(13)
③ 표준 용어를 논리 엔티티 속성으로 그대로 적용한다 (Korea Database Promotion Center, 2006).

4. 데이터 표준화의 구현 사례 및 검증

4.1 데이터 표준화의 구현 사례

본 장에서는 데이터 표준화를 구현한 사례를 중심으로 기술한다. 구축된 표준 및 모델 데이터 현황을 데이터 표준화가 수행된 프로젝트의 설계 단계 종료 시점까지 Table 2에 제시하였다.

Table 2. Status of standard and model data construction (2023. 03.16.)

Data classification	Detailed classification	Number of constructions
Standard Data	Standard Data	6,914
	Standard Domain	554
	Standard Terminology	36,547
	Standard Code	126,589
Model Data	Entity/Table	220
	Attribute/Column	3,221

Fig. 5는 표준 단어를 등록하는 화면이다. 영문 약어 및 단어 유형이 정의되어 각 표준 단어별로 있는 것을 알 수 있다.

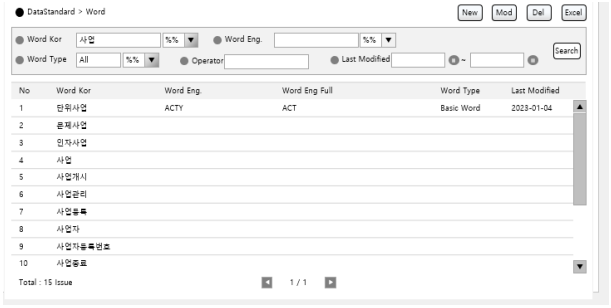

Fig. 5. Standard word registration.

Fig. 6의 화면은 표준 용어를 등록하는 화면이다. 영문명 및 도메인 결합되어 정의되어 있는 것을 각 표준 용어별로 알 수 있다.

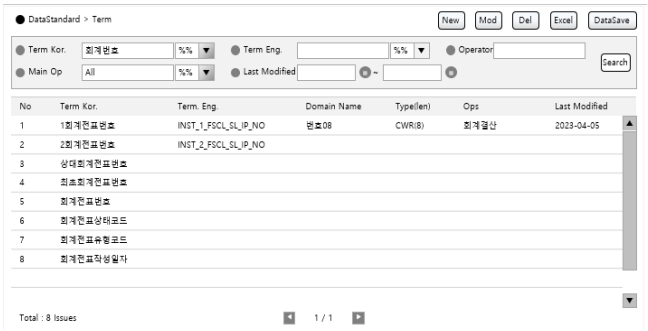

Fig. 6. Standard term registration.

Fig. 7은 표준 용어를 적용한 데이터모델 예이다. 우측은 물리 모델이고, 좌측은 논리 모델이다. 물리 칼럼은 논리 속성에 대응하는 표준 용어를 적용하여 별도 작업없이 자동 반영된다.

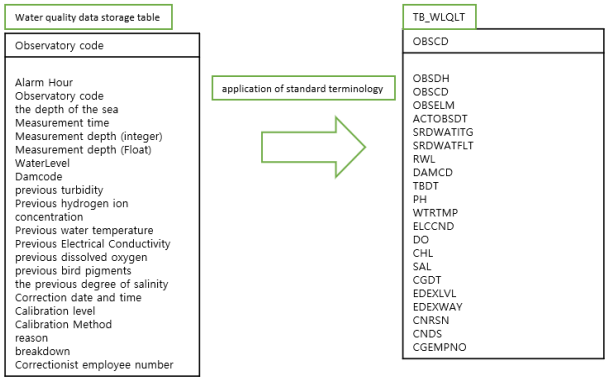

Fig. 7. Data Model.

4.2 데이터 표준화 효율성 검증

(1) 데이터 품질 측정 메트릭

오류가 발생한 비율을 이용하여 전체 데이터 구조의 일관성 측면에서 품질을 구하도록 정의한 메트릭을 품질 측정을 위한 메트릭으로 적용하였으며, 식(1)과 같다.

표준 데이터와 일관성이 맞지 않는 경우 오류로 각 컬럼별 오류 데이터(cij)는 판단하였으며, 가중치(wj)는 전체 칼럼 개수에서 각 컬럼의 개수가 차지하는 비율로 계산하였다.

<정의> 데이터 품질 측정 메트릭 Q

N : 총 품질 측정 대상 속성 개수

T : 칼럼별 가중치를 부여한 총 오류 속성 개수

Q : 데이터 품질 측정값

mi : 총 칼럼 수

cij : 각 칼럼에 해당하는 오류 속성의 개수

wj : 각 칼럼에 해당하는 가중치

$$Q = 1 - \frac{T}{N} \quad (식1) \qquad T = \sum_{j=1}^{m_i} c_{ij} \times w_j$$

(2) 표준화 정량적 효과

데이터 표준화가 수행된 설계 결과를 평가하기 위해 제시된 구현 방안을 적용하여 규모 및 일정이 유사한 데이터 표준화를 수행하지 않은 타 프로젝트와 식(1)을 적용하여 일관성에 대한 측정 결과를 비교하였다. Table 3에 제시된 결과에 따르면 본 논문의 구현 방안을 적용한 프로젝트가 그렇지 않은 경우보다 데이터의 구조적 품질 특성이일관성에 대한 우수함을 보여준다.

Table 3. Comparison between projects

Sortation	Case Project	Other Projects
Table Count	220	236
N	3,221	1,551
T	0	0.086
Q	0	0.086

Table 4는 데이터베이스 설계와 병행하여 데이터 표준화를 수행한 결과 설계가 진행되어 데이터의 구조적 품질 특성이 N값이 증가하더라도 거의 유사한 결과를 보여준다. 이는 표준 데이터의 확보가 선행되지 않은 경우에도 데이터 베이스설계와 표준데이터의 구축을 병행하는 경우 데이터품질 수준의 적정성이 확보됨을 나타낸다(Yang and Choi, 2003).

Table 4. Comparison by stage

Sortation	January	February	March
Table Count	109	186	220
N	2036	3129	3221
T	0	2	0
Q	0	0	0

(4) 표준화 정성적 효과

데이터베이스 설계를 수행한 결과 위 구현 방안을 적용하여 정량적 효과인 다음과 같은 정성적 효과를 구조적 품질 향상 측면 외에 갖게 되었다.

① 표준화된 테이블 설계가 자동적으로 수행되어 논리 데이터 모델링과 동시에 설계 생산성 향상

② DB 구조의 변경에 대한 영향도를 쉽게 파악할 수 있게 표준화된 칼럼을 통하여 운영 편의성 제공

③ 프로그램의 용어 표준에 활용하여 구축된 표준용어 사전으로 개발 표준화를 유도하고, 데이터 컨버전의 기초 자료 활용

④ 표준화된 의사소통 수단을 통해 EUC(End-User-Computing) 및 데이터의 연계, 통합을 실현할 때 관리의 편의성 제공

4.3 데이터 품질 일관성 검증

(1) 데이터 품질 지수

데이터 품질 지수는 연구에서 관리되고 있는 데이터의 정합성에 대한 정량적 종합 지표와 업무규칙 측정 결과의 정합성의 대푯값을 의미한다. 품질 현황을 측정하기 위해 작성되는 데이터 품질 지수는 정보시스템의 품질 추이 분석이나 목표 수립에 필요한 기초 자료로 활용된다. 데이터 품질 지수를 계산하기 위해서는 데이터 품질 기준, 기준별 가중치 그리고 산술식이 필요하다.

(2) 상세 데이터 품질 기준

Table 5. Detailed data quality criteria

Quality Criteria	Application Criteria	Meaning	Derived Details Quality Standards
accuracy	up-to-date maintenance	Information generation, collection, and update cycles should be maintained.	up-to-dateness
	business relationship	Column values related to business relationships must satisfy the relevant business rules.	Accuracy of business rules
	correlation	If the columns are semantically related to each other, the precedence relationship and the calculated/aggregated values must be exactly the same.	Accuracy of predecessor relationship
			Calculation/ counting accuracy
consistency	Code Match	The interrelationships of data using the integrated code table should be maintained.	Reference Code Consistency
	Relationship	Reference integrity between tables shall be observed.	Reference Integrity
	Data Flow	If data is moved by generating or processing data, all related data must match.	Data flow consistency
	Overlapping	When duplicate columns are randomly generated and utilized for management purposes, duplicate column values must match.	Column Consistency
uniqueness	non-redundancy condition	The value of the column must be unique and must not be duplicated.	Identifier uniqueness
			Combination key uniqueness
perfection	Missing condition	The value of the required data column shall be free from omission.	Individual Completeness
			conditional completeness
effectiveness	Domain Type	The value of the column must meet the specified data range and domain.	Date Validity
			Type Validity
			Range Validity

(3) HNS 데이터 수집시 중요도 산정

중요도를 산정하여 가중치를 정의하는 방법은 임의(Adhoc) 방법과 사전 정의(Predefined) 방법이 있다.

사전정의 방법은 품질진단 대상, 데이터베이스 구축, 품질진단 목적, 품질진단 활용 목적 등을 고려하여 중요도를 산정하는 방법이다. 또한, 사전정의 방법은 관련 전문가의 의

견도 중요도를 산정하여 반영한다.

임의 방법은 사전에 정의된 중요도가 없어 별도의 중요도를 산정하는 경우에 통계적 분석 방법을 통해 가중치를 정량적으로 산출하는 방법이다. 임의 방식은 품질관리자, 품질담당자, 업무 담당자 등을 포함한 관련 담당 전문가들의 의견을 반드시 포함해야 한다. 계층적 분석법(AHP, Analytic Hierar chy Process)은 대표적인 임의 방법이다.

최종 가중치는 2개의 의사결정 기준에 대해 상대적 가중치를 결정한 후에 각 가중치에 로우별 평균을 곱하여 산정된다. 앞의 표에서 중요도와 가중치 비율을 0.4 : 0.6으로 가정하여 최종 가중치 산정 결과는 다음 Table 6과 같이 나타난다.

Table 6. Raw data based on importance

compare reference	index1	index2	index3	index4	index5
index1	0.09523	0.08163	0.11111	0.0625	0.15384
index2	0.28571	0.24489	0.22222	0.25	0.30769
index3	0.38095	0.48979	0.4444444	0.5	0.30769
index4	0.19047	0.12244	0.11111	0.125	0.15384
index5	0.04761	0.06122	0.11111	0.0625	0.07692

Table 7. Raw data based on priority

compare reference	index1	index2	index3	index4	index5
index1	0.1111	0.0625	0.041619	0.0625	0.0769
index2	0.2222	0.232	0.209714	0.2992	0.5028
index3	0.4444	0.44	0.390952	0.5	0.2936
index4	0.1111	0.133	0.157	0.118	0.12
index5	0.1111	0.0785	0.091238	0.0625	0,1234

Table 8. Final weight selection for each indicator

Average Weight / index	Average of ROWs based on importance(①)	ROW average based on priority(②)	final weight (①×0.4)+(②×0.6)	
index1	0.101	0.070	0.083	≈ 0.1
index2	0.262	0.294	0.281	≈ 0.3
index3	0.425	0.413	0.418	≈ 0.4
index4	0.141	0.127	0.133	≈ 0.1
index5	0.072	0.095	0.085	≈ 0.1

계층적 분석법에서 중요한 것은 지표 간의 비교평가에서 논리적 일관성을 유지하는 것이다. 이를 위해 지표 간 비교평가와 관련된 전문가의 의견을 반영해야 한다. 이 때, 가능하다면 다수의 전문가의 의견을 반영하는 것이 바람직하다.

지표 간 비교평가에서 논리적 일관성은 일관성 지수(CI, Consistency Index)로 확인할 수 있다. 이때, 응답자가 논리적 모순을 야기하면 일관성 지수가 증가하여 응답자의 신뢰성이 떨어진다고 판단한다. 이러한 이유로 일관성 지수는 응답에 대한 논리적 모순을 검증하는 것이 필요하다. 다음은 일관성 지수를 계산하는 수식이다.

$$\text{일관성 지수 CI} = \frac{\lambda\max - n}{n-1}$$

여기서 $\lambda\max$는 최대 고유값(Principal Eigen Value)이고 n은 행렬의 차원이다. n x n 정방행렬 [A]와 n x 1 가중치 행렬 [W]를 곱하면 신규 n x 1 가중벡터 행렬 [Y]가 산정되는데, 가중 벡터행렬의 구성요소 Y1 … Yn과 가중치 W1…Wn을 이용하여 다음과 같이 $\lambda\max$를 계산한다.

[A] X [W] =[Y]일 때,

$$\lambda\max = (Y1/W1+Y2/W2+ \cdots +Yn/Wn)/n$$

이와 같은 계산 방법에 따라 중요도와 가중치 계산에 대한 일관성 지수를 산정해 보면 앞의 표는 모두 '0.1'보다 작기 때문에 비교평가에 대한 논리적 일관성이 유지되고 있어, 최종 가중치 산정 결과는 신뢰성이 있다고 판단한다.

계층적 분석법을 사용할 때 주의할 점은 비교항목이 너무 많으면 복잡성이 증가하여 적용이 쉽지 않다. 일반적으로 비교 항목이 15개를 초과하면 최종 가중치 산정이 불가능하다. 따라서, 계층적 분석법을 통해 가중치를 산정할 때는 중요도 선정 항목의 세부 유형 항목을 비교항목으로 선택하여 가중치를 선정하는 것이 바람직하다(Korea Database Promotion Center, 2021).

5. 결 론

HNS 데이터의 품질 확보를 위한 중요한 영향 요인으로 데이터 품질은 데이터 표준화를 통해 데이터의 구조적 품질 수준을 보장할 필요가 있다.

HNS 데이터 표준화 활동들이 수행되면 사용자들은 정확한 데이터를 사용할수 있고, 올바른 의사결정을 내릴수 있다. 경쟁력ㄷ 확보에 많은 영향을 미친다. 명칭의 통일로 인한 명확한 의사소통의 증대하고, 필용한 데이터의 소재 파악에 소요되는 시간 및 노력 감소하며, 필요한 데이터 형식 및 규칙의 적용으로 인한 데이터 품질 향상으로 정보시스템 간 데이터 인터페이스 시 데이터 변환, 정제 비용 감소를 기대한다.

실무에서 HNS 데이터베이스 설계와 병행하여 데이터 표준화 개념을 수행할 수 있게 함으로써 설계 생산성과 데이터의 구조적 품질을 향상시킬 수 있도록 하는 방안을 제시하였다. 구현 방안으로는 관리 프로세스 및 설계지침과 데이터 표준화를 위한 환경을 기술하였다. 본 논문에서 제시한 구현 방안의 효율성을 측정하기 위해 DB 속성 설계의 일관성에 대한 메트릭을 적용하였으며, 그 결과 데이터 표준화를 수행하지 않는 사례에 비해 일관성 측면에서 데이터의 구조적 품질 특성이 우수한 결과를 보였다.

본 논문에서 제시된 구현 방안은 HNS 표준 데이터 구축이 선행되지 않은 경우에도 구조적 품질 수준이 보장된 데이터베이스 설계를 수행하고자 하는 실무에 기여할 수 있다.

후 기

이 논문은 2023년도 해양수산부 재원으로 해양수산과학기술진흥원의 지원을 받아 수행된 연구이다(20210660, 해양위험유해물질(HNS) 배출 등 관리기술 개발사업, 해양산업시설 배출 위험유해물질 영향평가 및 관리기술 개발).

This research was supported by Korea Institute of Marine Science & Technology Promotion(KIMST) funded by the Ministry of Oceans and Fisheries, Korea(20210660).

References

[1] DoD4000.25-13-M.(1996), DoD Logistics Dsata Eleme nt Stadardization and Managmetn Program Procedures, 6.

[2] English, L. P.(1999), Impoving Data Warehouse and Business Information Quality.

[3] IOS/IEC Standard(2000), ISO/IEC 9126.

[4] Kim, C. S. and J. S. Park(2004), Development of maturity models through data structural analysis.

[5] Korea Database Promotion Center(2006), Data Quality Management Guidelines Ver 2.1

[6] Korea Database Promotion Center(2017), Data Quality Guidelines, 6

[7] Korea Database Promotion Center(2021), Data Quality Guidelines, 12

[8] Kulpa, M. K. and K. A. Johnson(2003) Interpreting the CMMI, AUERBAH, 6

[9] Strong, D. M., Y. W. Lee, and R. Y. Wang(1997), Data Quality in Context, Communcations of the ACM, Vol. 40. No. 5.

[10] Yang, J. Y. and B. J. Choi(2003), Data Quality Measurement Tools, Journal of the Korean Society of Information Science, Computing Reality, Vol. 9, 3.

Received : 2023. 09. 20.
Revised : 2023. 10. 26.
Accepted : 2023. 10. 27.