

Geospatial Data의 품질평가를 위한 Framework A Framework for Quality Evaluation of Geospatial Data

조 기 성*
Cho, Gi Sung

要 旨

최근 날로 정보기술이 발전하고 사회가 다양화되면서 다양한 데이터를 공유하기 위한 데이터표준화에 대한 요구가 증대되고 있어 이와 같은 표준화를 위해 우선적으로 지형공간데이터의 정확성과 신뢰도를 나타내는 데이터품질의 정의 및 평가과정에 관한 연구가 요망되고 있다.

본 연구에서는 세계 각 국에서 채택하고 있는 대표적인 데이터 품질요소의 정의 및 평가방법을 비교고찰하여 다음과 같은 결과를 얻었다.

첫째, 우리 나라의 국가지형공간데이터 교환포맷의 표준내에 데이터의 품질을 평가하기 위한 요소의 정의는 평가기준이 명확한 ISO/TC211의 안을 수용하는 것이 바람직하다고 판단된다.

둘째, 본 연구를 통하여 품질평가요소에 따른 보다 합리적인 지형공간정보 데이터의 품질평가과정을 제시하였으며, 이를 KSDTS에도 적용할 것과 데이터 품질평가결과의 보고형식 등을 더욱 더 명확히 하여 국립지리원의 수치지도작성규칙등에도 포함시키는 것을 제안하고자 한다.

셋째, 다양한 sampling방법에 관한 연구 및 우리 나라에 적합한 품질허용기준(AQL)의 정립과 대량의 데이터 품질을 신속하게 자동평가할 수 있는 컴퓨터 프로그램 작성 등이 요망된다.

Abstract

Lately, the demand for data standardization become increased to obtain various data jointly along with development of information technology and diversity of society. Thus the research on the definition and evaluation of data quality indicating accuracy and confidence of geospatial data, is required for this standardization.

In this study, by virtue of comparison of definitions and evaluation methods of data quality element being selected from representative countries, the following results were obtained:

* 정희원, 전북대학교 공과대학 토목 및 환경공학부 조교수

(1) Application of ISO/TC211's Draft having accepted evaluation standard to KSDTS(Korea Spatial Data Transfer Standard) is desirable for definitions of data quality elements.

(2) This study presented the quality evaluation of much more reasonable geospatial data accompanying with quality element. Furthermore, this study suggests that this evaluation be applicable to KSDTS and be contained in the digital map product specification of National Geography Institute with more clearness of a report form of data quality evaluation result.

(3) Studies on various sampling methods, establishment of AQL(Acceptable Quality Level) suitable for our country, and computer programming which can rapidly and automatically evaluate mass much of data are required.

1. 서론

최근 컴퓨터의 고성능화, 저가격화 및 인터넷 등의 정보통신망(network)이 급속하게 발전하고 있어 정보에 대한 보다 높은 처리와 유통을 위한 환경이 정비되고 있다. 컴퓨터와 정보통신망의 정비는 고도화사회를 지탱하기 위해 불가결한 것으로 그 중에서도 데이터가 여러 목적에 따라 세분화되고 특히 공간데이터의 필요성을 인식하고 자신들의 어플리케이션에 적합하도록 독자적으로 데이터를 구축해 나감에 따라 데이터의 양이 급속히 증가하고 있기 때문에 정보자원의 정비 및 그 정보의 품질을 유지하는 것 또한 중요한 요소 중에 하나가 되고 있다. 그리고 지형공간정보는 이를 이용하는 각 기관 별로 자체적으로 제작하였거나 추진할 계획을 가지고 있어 호환성이 결여될 우려가 있는 등 앞으로 여러 가지 문제점을 내포하고 있어 이들 지형공간정보에 대한 표준 및 질적 평가 기준을 확립해야 할 필요성이 높아지고 있다. 따라서 세계 각국에서는 이에 대한 여러 가지 연구 및 운용이 이루어 지고 있으며 미국에서는 국가지리공간데이터기반(National Spatial Data Infrastructure: NSDI) 프로젝트에 따라 정부, 주, 지방자치단체에 의한 지형공간데이터의 구축 뿐 만 아니라 처리, 배포, 이용을 위한 기술, 정책, 표준화 등을 포함한 효율적이고

종합적인 관리를 위해 노력하고 있으며, 정부기관이 보유하고 있는 데이터의 소재정보 등을 인터넷을 통하여 제공하기 위해 연방지리데이터위원회(Federal Geographic Data Committee)에 의해 metadata표준이 제정되어 있다. 그리고 USGS(U.S. Geological Survey)가 주도한 SDTS(Spatial Data Transfer Standard)는 정부, 기업, 학계 전반이 9년동안의 개발 과정을 거쳐 1992년 7월 29일에 미국연방정보처리기준(FIPS Publication 173)으로 승인된 것으로 호주, 뉴질랜드와 더불어 우리 나라가 국가표준으로 정한 공간데이터교환표준이 있다.

일본에서는 1993년 프랑스로부터 측량, 지도관계의 국제표준을 작성하는 전문위원회의 설립제안을 받은 이래 국립지리원의 1994년 지리정보 디렉토리 데이터베이스(GIDD: Geographical Information Directory Database)의 구축에 관한 연구를 시작으로 건설성의 GIS연구회, 정부주도의 GIS관계 성, 청 연합회의등 지형공간데이터의 효율적인 정비 및 그 상호이용의 촉진에 노력하고 있으며, ISO(International Standard Organization)의 기술위원회의 하나인 TC211 등의 지형공간정보 표준화 국제전문위원회에 적극 참여하고 있다.

유럽에서는 유럽표준화기구(Comite Europeen de Normalisation: CEN)의 TC287이라는 소위원회를 통하여 지형공간정보의 표준화를 진행시키고 있으며,

NATO국가들의 군지도제작조직들에 의해 구성된 DGIWG(Digital Geographic Information Working Group)는 공간데이터의 교환표준으로 DIGEST를 채택, 공표하였다.

각국의 수로국업무를 총괄하는 국제기구인 IHO(International Hydrographic Organization)는 S57(or DX-90)을 교환표준으로 하고 있다.

ISO/TC211은 1994년 6월에 구성된 ISO산하 지형공간 정보분야의 표준화를 위한 기술위원회로서 24개국의 정회원국과 13개의 옵서버국으로 구성되어 있으며 우리나라는 투표권을 행사할 수 있는 정회원국이다. 이는 5개의 소그룹으로 나누어서 공간데이터의 모델, 관리, 품질, 서비스, 프로파일, 코드화, 연산자 등의 정의 및 공간데이터의 표준화를 위한 작업을 1998년을 목표로 진행하고 있다.

국내에서도 1994년 5월 경제장관회의에 보고된 국가 지리정보체계(NGIS) 구축방안을 기초로한 1995년 5월 19일 NGIS 추진위원회를 통하여 국가지리정보체계구축사업 기본계획을 심의 확정하고 국가지리정보체계제반사업을 범부처적으로 적극 추진중에 있다. NGIS의 여러분과중 표준화분과에서는 NGIS 구축에 필요한 공간정보의 표준화 방안, 전산망 표준화지침에 의거 GIS표준설정, ISO 등 국제적인 표준과의 상호교류를 추진하고 있다. 실제적으로 1995년 12월 18일 제5차 공통데이터 교환포맷 소위원회의 회의 결과 국가데이터교환포맷 표준으로서 SDTS를 선정하고 한국실정에 맞고 미흡한 부분을 보완하여 한국형공간데이터 교환표준(KSDTS)을 작성하기 위해 노력하고 있다.

이와 같이 지형공간정보데이터의 표준에 대하여 우리나라 뿐만 아니라 세계 각국이 노력하고 있는 것은 정보기술이 발전하고 사회가 다원화되면서 다양한 데이터를 다수가 공동으로 필요하게 되어 데이터를 공유할 수 있는 표준화에 대한 요구가 증대되기 때문이며, 각 국가별 또는 국제적인 표준화가 이루어 질 경우 각 시스템간의 연동성 및 상호운영성 확보를 가능케 하고 개별적인 데이터구축으로 인한 재정적인 중복투자 방지에도 크게 기여하기 때문이다.

따라서 이와 같은 지형공간정보의 표준화의 근본적인 목적의 하나인 데이터의 공유와 활용을 위해서는 이

미 구축되어 있는 또는 앞으로 구축할 데이터의 정확성 및 품질에 관한 신뢰도 등에 관한 자료가 요구된다. 이 데이터의 품질에 관한 정보는 데이터의 가치로서 뿐만 아니라 데이터이용에 앞서 선택과정에서 절대적으로 필요하게 되었다. 데이터 생산자에게 있어서는 생산을 더욱 확실하게 할 수 있고, 데이터 사용자는 앞으로 다양한 수준의 데이터 품질을 접하게 될 것이며, 사용자에 따라 높은 정확도를 필요로 하는 경우도 있으며, 보다 덜 정확한 데이터를 요구하는 경우도 있게 되는 바와 같이 특수목적에 사용될 데이터의 적합성을 판단하는데 기준이 될 것이다. 그러므로 본 연구는 지금까지의 세계 각국에서 채택하고 있는 지형공간데이터의 표준중에서 현재 각 소위원회별로 제정하고 있는 국제표준인 ISO/TC211을 기준으로하여 우리나라에서 한국형공동데이터교환표준을 개발하는데 원칙으로 하기로한 SDTS, 영국의 국가지도 제작기관인 Ordnance Survey와 민간부문의 공동노력으로 이루어진 NJUG(National Joint Utilities Group)등 본 연구 목적에 적합한 데이터품질 요소 및 평가방법이 비교적 상세히 기술된 표준을 대상으로 하여 지형공간데이터의 품질의 요소를 비교분석하고 사용자가 응용 또는 요구에 맞는 데이터를 선택하는 데 데이터 품질이 유용될 수 있도록 하는 데이터품질의 평가법 및 과정 그리고 추후연구과제를 제시하되로서 한국형 공간데이터 교환표준개발에 보탬이 되고자 하는데 목적을 두고 있다.

2. 지형공간데이터의 품질평가의 정의 및 요소

2.1 데이터 품질평가의 정의

지형공간정보의 품질이란 데이터에 나타낸 사상이나 속성이 추상적인 실세계를 얼마나 잘 묘사 하였는가의 정도를 나타내는 것으로서 데이터생산자가 추상적인 세계를 나타내는데는 일반적으로 definite product spec이나 project description을 사용한다. 또한 이용자 측면에서의 데이터품질은 그들의 요구에 얼마나 잘맞고 이용하고자 하는 응용분야나 이용적합도에 어느

정도 접근해 있는 지에 깊은 관계가 있다. 이를 위해서는 지형공간데이터의 품질을 일관성있고 표준화된 측정방법에 의해 평가하여야 한다. 이 품질 평가법은 데이터 품질면에서의 모든 것을 포함시키기는 어렵고 데이터 생산자와 이용자에게 참고사항을 제시하는 것이다. 데이터 품질에 관한 요소의 정의를 기초로하여 작성된 데이터 품질에 대한 평가과정을 거치므로써 데이터 생산자는 그들이 생산한 데이터 품질을 자체평가 할 수 있고, 이용자는 주어진 데이터에 대하여 자기의 요구에 맞는 지를 점검할 수도 있다.

그러나 지형공간데이터의 내용과 생산환경이 자주 변화될 수 있기 때문에 표본추출(sampling)과 같은 품질 평가과정의 일부는 특수조건, 데이터 생산환경, 사용법 등에 따라 발생하는 요구조건에 맞게 수정되고 변경될 수도 있으나 이와 같은 경우에도 품질평가과정의 시스템에 따라 평가를 실시하면 일관되고 표준화된 결과를 얻을 수 있을 것이다.

2.2 데이터의 품질요소

데이터의 품질을 나타내고 평가하기 위한 요소에는 크게 직접적인 요소와 개념적인 요소로서 나눌 수 있으며, 직접적인 품질요소는 다음과 같이 완전성(completeness), 논리적 정합성(logical consistency), 위치정확도(positional accuracy), 주제정확도(thematic accuracy) 그리고 시간적인 정확도(temporal accuracy)등으로 나눌 수 있다. 개념적인 요소는 데이터의 특징을 나타내는 요소로서 metadata에 포함되는 데이터의 목적, 연역, 사용법등이다.

2.2.1 완전성

완전성란 실세계의 모든 현상과 속성이 product spec(예:수치지도작성규칙)에 따라 속성값을 가진 사상으로서 기록(encode)되었는가의 정도를 나타낸다. 이 완전성은 사상의 완전성(feature completeness)와 속성의 완전성(attribute completeness)으로 분류할 수 있다.

사상의 완전성이란 모든 실세계의 현상이 product spec에 따라 omission과 comission형태로 내적 또는 외적인 사상으로 정의되는 정도를 나타낸다. 속성의 완전성은 속성, 사상의 속성type, 사상type 등이 product spec에 따라 omission과 comission형태로 내적 또는 외적으로 정의되는 정도를 나타낸다.

2.2.2 논리적 정합성

논리적 정합성은 데이터 구조에 대한 규칙과 내부구조(internal structure)형태인 product spec과 어느 정도 일치하는가와 실세계 현상의 속성이 product spec에 따라 정확히 기록되어 있는가의 정도를 나타낸다. 이 논리적 정합성은 위상정합성(topologic consistency), 속성정합성(attribute consistency), 형식정합성(formal consistency)으로 나눌 수 있다.

위상정합성은 product spec에 의해 지정된 사상들 간의 위상관계에 있어서 논리적인 법칙이 지켜졌는가의 정도를 나타낸다. product spec에 지정된 가장 일반적인 위상관계의 예로서 평면그래프의 위상이나 사상의 포함관계(inclusion relation)등과 같은 순위적 관계(ordered relation)등이다.

속성정합성은 속성값이나 type이 product spec에 나와 있는 논리적 법칙(logical rule)에 어느 정도 일치하는가의 정도를 나타내는 것으로서, 예를 들면 어느 속성값의 "domain"과 또다른 속성들과의 사이에 존재하는 함수관계 등의 일치를 말한다.

형식정합성은 데이터를 이루고 있는 구조자체가 product spec에 맞게 구성되어 있는지의 정도를 나타낸다.

2.2.3 위치 정확도

위치정확도는 공간실체 사실(entity instance)의 위치에 대한 데이터의 공간실체 객체(entity objects)위치값의 정확도 또는 product spec에 의해 생성된 외적인 참조데이터의 정확도로서 정의된다. 이 위치정확도는 절대위치정확도와 상대위치정확도로서 대별된다. 절대위치정확도는 product spec에 따라 지구상의 위치

를 나타내는데 사용된 표준기준계(standard reference system)에 대한 위치정확도이다. 절대 수평위치 정확도는 사상과 사상type의 위치적인 값과 표준기준계에 관련된 추상적인 사상의 위치와의 거리로서 정의할 수 있다. 절대 높이정확도는 사상과 사상type의 높이값과 표준기준계에 관련된 추상적인 사상의 높이와의 연직거리로서 정의할 수 있다.

상대 위치정확도는 product spec에 따라 작성된 데이터의 다른사상이나 국지적인 기준계에 대한 사상의 상대적인 정확도를 나타낸다. 상대 수평위치정확도는 사상과 사상type의 위치적인 값과 국지적인 기준계에 관련된 추상적인 사상의 위치와의 거리로서 정의할 수 있다. 거리는 절대 수평위치와 마찬가지로 정의된다. 상대 높이정확도는 사상과 사상type의 높이값과 국지적인 기준계에 관련된 추상적인 사상의 높이와의 연직거리로서 정의할 수 있다.

2.2.4 주제 정확도

주제정확도(thematic accuracy)는 데이터의 비위치적인 특성이나 비 시간적인 특성의 정확도를 나타낸다. 주제정확도는 연속적인 변수정확도(continuous value accuracy), 서수적인 변수정확도(ordinal value accuracy), 공칭적인 변수정확도(nominal value accuracy)로 나눌 수 있다.

각 변수 정확도는 product spec에 의해 작성된 사상의 각각 변수들의 특징들 사이의 차이(discrepancy)로서 정의된다.

2.2.5 시간적인 정확도

시간적 정확도란 시간적인 특성의 정확도로서 즉 데이터의 관측, 입력, 출력, 삭제, 검증 등의 시간 및 시기의 측정정확도를 나타낸다.

데이터 생산자로부터 이미 제공된 데이터의 품질은 다음 두 가지 조건에 영향을 받게 된다. 하나는 데이터를 수정하거나 제거, 추가한 경우이고, 또 하나는 데이터를 생성하기 위한 product spec이 변경된 경우이다.

많은 데이터는 고정되어 있지 않기 때문에 데이터를 수정하는 경우가 자주 빈번히 있게 된다. 이와 같이 변경, 수정된 데이터의 경우에는 데이터 생산자가 품질을 재평가하여야 한다.

그리고 product spec의 변화는 초기데이터 구축과정에서 자주 발생되며 데이터 품질정보를 얻기 이전에 생기게 된다. 데이터 product spec이 변하게 되면 데이터의 품질 또한 당연히 변하기 때문에 생산자는 최근에 사용되고 있는 최신product spec에 따라 최신데이터의 품질을 평가하여야 한다.

2.3 데이터 품질의 개념적 요소

데이터 품질평가의 개념적 요소는 데이터의 특성에 관한 데이터로서 meta데이터의 내용에 포함되는 것으로서 데이터의 목적, 연혁, 사용법등이다. 이 데이터는 이용자에게 자기의 기준을 이용해서 어떤 특수한 응용에 있어서 데이터의 이용적합도를 판단하는 데 이용될 수 있다. 또한 이 개념적 요소는 품질을 수량적으로 측정치 않는다는 점에서 데이터 품질요소와 구별된다. 즉 질적 개념의 품질요소라 할 수 있다.

데이터 목적에 관한 정보에는 데이터 생산자의 데이터 생성의 근본적인 목적을 기술하며 데이터의 이용분야 및 의도가 포함된다.

데이터 연혁에 관한 정보는 데이터의 수집 및 획득에서 부터 편집, 처리등 지금의 상태에 이르기까지의 유도과정을 기술하고, 특히 자료원(source data)에 관한 정보를 상세히 기술하고 처리과정 또는 연역정보에 좌표 및 기타 변환관계나 특별한 이벤트에 대한 기록도 꼭 포함되어야 한다.

데이터 사용법에 관한 정보는 데이터 이용에 필요한 여러 정보 및 사용법이 생산자로부터 이용자에게 명확히 전달되도록 기술하여야 한다.

2.4 각표준의 데이터 품질평가요소의 비교고찰

지금까지 세계 각국에서 채택하고 있는 지형공간데이터의 표준중에서 우리 나라에서 한국형 공동데이터

Table 1. 각 표준의 데이터 품질평가요소

각 표준	SDTS	ISO/TC211	Ordnance Survey
지형공간데이터의 품질평가요소	1.데이터연혁 (data lineage)	1.완전성(completeness) 사상 완전성 속성 완전성	1.형식 정합성 데이터 포맷 데이터의 수 feature code 정확도
	2.위치정확도 (positional accuracy)	2.논리적 정합성 (logical consistency) 위상 정합성 속성 정합성 형식 정합성	2.위치정확도
	3.속성정확도 (attribute accuracy)	3.위치정확도 (positional accuracy) 절대위치 정확도 상대위치 정확도	3.논리적 정합성 면적위상 정합성 선연결 정합성
	4.논리적정합성 (logical consistency)	4.주제정확도 (thematic accuracy) 연속적인 변수 서수적인 변수 공칭적인 변수	4.text
	5.완전성 (completeness)	5.시간적인 정확도 6.개념적요소(metadata)	5.completeness

교환표준을 개발하는데 원칙으로 하기로한 SDTS와 현재 각 소위원회별로 제정하고 있는 국제표준인 ISO/TC211, 영국의 Ordnance Survey와 민간부문의 공동노력으로 이루어진 NJUG의 지형공간 데이터의 품질요소를 비교하여 Table 1에 나타내었다.

Table 1에서 알 수 있듯이 SDTS에서 정의한 품질요소는 metadata에 해당하는 데이터연혁과 위치정확도, 속성정확도, 논리적정합성, 완전성등 외관상으로는 어느 정도 거의 모든 품질요소를 포함하고 있으나 실제 내용에 있어서는 각 요소에 대하여 명확히 규정되어 있지 않다. 한편 Ordnance Survey는 처음 데이터를 접할 경우 데이터의 형식이 product spec의 형식에 맞는 지를 검사하는 형식정합성의 평가과정이 포함되어

있으며 위치정확도, 논리적정합성, metadata의 내용에 해당하는 text와 완전성을 품질평가요소로 하고 있다. 그리고 ISO/TC211, Working Group3, No13의 Draft에 나타나 있는 지형공간품질요소의 정의는 2.2절에 나타난 바와 같이 완전성, 논리적 정합성, 위치정확도, 주제정확도, 시간적인 정확도, metadata에 해당하는 개념적 요소에 이르기까지 세부적으로 제시하고 있다. 물론 세부적이라 하여 완벽한 것은 아니지만 데이터의 품질을 평가하기 위한 요소의 정의를 정확히 제시 하므로써 앞으로의 컴퓨터를 이용한 자동 품질평가 프로그램작성 등에 도움을 주기 위해서는 기준이 명확히 제시되어 있는 ISO/TC211의 안이 지형공간데이터의 품질평가요소로서 적합하다고 판단된다.

따라서 우리 나라의 국가지형공간데이터 교환포맷 (KSDTS)의 표준내에 데이터의 품질을 평가하기 위한 평가요소는 SDTS를 그대로 사용하는 것 보다는 ISO/TC211의 안을 적극 수용하여 데이터의 품질에 관하여 이용자에게 뿐 만 아니라 데이터생산자에게도 도움이 될 수 있도록 기준을 명확히 제시하는 것이 바람직하다고 판단된다.

3. 지형공간데이터 품질평가의 Framework

3.1 데이터 품질평가 과정

지형공간 데이터의 품질평가과정의 일반적인 흐름도는 Fig 1과 같은 과정으로 구성되어 있다.

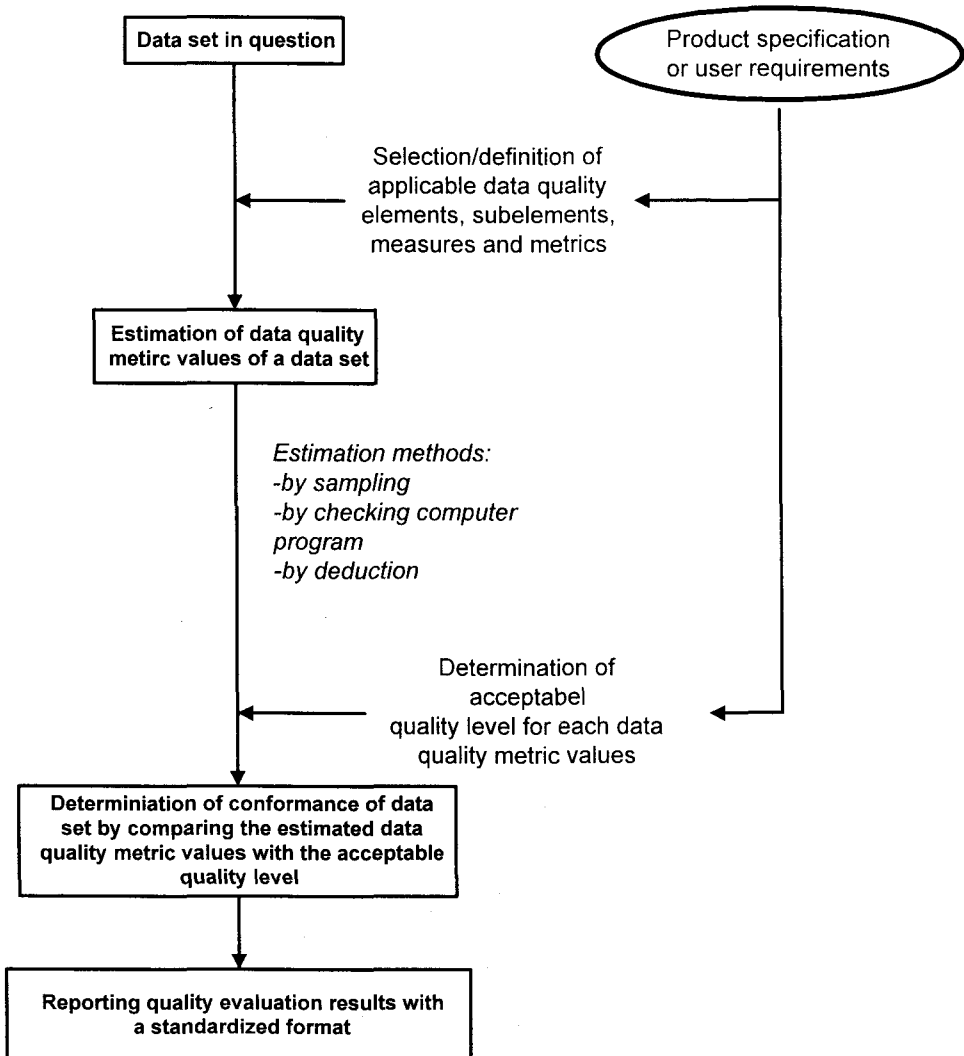


Fig 1. Framework flow of quality evaluation procedure for geographic data

우선 데이터가 product spec에 의해 지정된 데이터 형식과 일치하는지 아닌지를 점검하고 데이터에 오차가 없을 경우에만 품질측정에 관한 점검을 실시한다. 데이터의 품질을 결정하는 측정값을 추정하는 과정은 세 가지의 접근 방법이 있다. 첫째는 표본추출에 의한 방법으로 데이터로부터 추출한 표본(sample)으로부터 품질을 평가하기 위한 값을 계산한다. 둘째는 자동

조사법으로서 컴퓨터 프로그램에 의해 데이터의 모든 항목(item)을 점검하여 품질을 평가하는 방법이다. 셋째는 연역적 방법으로서 데이터의 품질측정값을 외부 정보(external knowledge)로부터 추정하는 방법이다. 즉 본래데이터의 품질에 관한 데이터와 생산과정에서의 오차전파모델 등의 자료로부터 추정하는 방법이다.

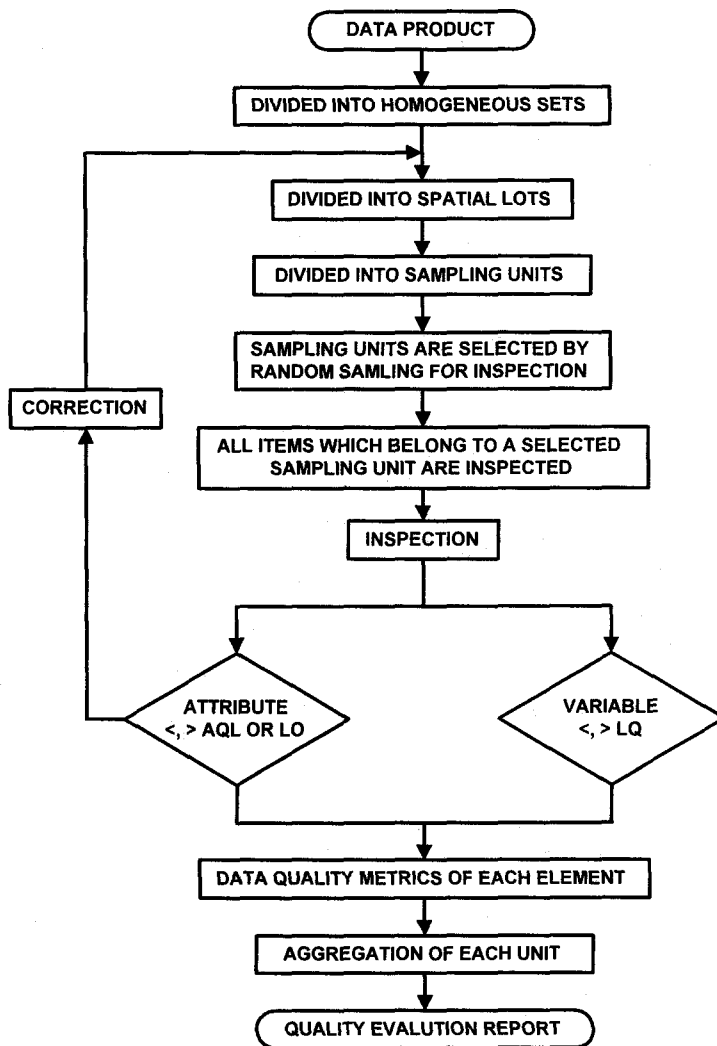


Fig 2. Flow chart of inspection by sampling

다음 과정은 계산된 품질측정값을 인정할 수 있는 품질 수준(Acceptable Quality Level: AQL)과 비교하므로 데이터의 적합성을 결정하고 다음은 마지막 과정으로서 위의 품질측정과정으로 부터 얻어진 결과를 주어진 형식에 따라 품질평가보고서로서 작성하여 사용자에게 제공하는 것이다.

3.2 각 품질요소에 따른 품질측정법

데이터에 대한 품질측정의 평가방법은 크게 두 가지로 나눌 수 있으며 그 하나는 데이터가 product spec에 따르는 지 어떤 지를 결정하기 위해 표본추출에 의해 계산한 값을 측정값으로 하거나 컴퓨터 프로그램에 의해 모든 데이터를 점검하여 품질을 평가하는 직접측정법이다. 다른 하나는 연역적, 간접적 측정법으로서 추론(deduction)에 의해 데이터의 적합도나 측정값을 얻는 방법이다. 이와 같은 추론에 이용될 데이터의 예는 데이터 생성에 사용된 본래데이터의 품질에 관한 정보나 생산과정의 오차전파모델등으로서 데이터생산자에게 요구되고 있는 ISO9000규정에 의한 품질의 기록이나 설명서가 추론을 위한 좋은 정보가 될 수 있다.

3.2.1 직접적인 방법

1. 컴퓨터 프로그램에 의한 자동점검과정

컴퓨터 프로그램에 의한 자동품질평가과정은 품질측정계산에 있어서 데이터의 내적정보 즉 파일포맷점검과 비적합성 항목의 퍼센트나 데이터의 품질평가결과의 평균치 등의 계산도 컴퓨터 프로그램이 이용가능하며 가장 일반적인 예로서 product spec에서 지정된 화일형식(file format)의 일치도 점검등이다. 이와 같은 자동화과정에서 사용된 점검프로그램의 이름, 알고리즘, 이용가능한 다른 참고정보등은 품질에 관한 신뢰성을 입증하기 위해 품질평가보고서에 상세히 기록하여야 한다.

2. sampling을 기초한 추정과 평가과정

표본추출에 의한 일반적인 생산품에 관한 조사는 다음과 같은 ISO2859(sampling procedures for

inspection by attribute), ISO3951(sampling procedures and charts for inspection by variables for percent non-conforming)에 규정되어 있으나 지형공간데이터의 특성을 고려할 때 다음과 같은 사항을 수정하여 적용하여야 한다. omission은 지형공간데이터 품질 측정을 나타내는 방법중의 하나로서 omission을 점검하려면 표본추출은 데이터가 아니라 실 세계의 사상의 type, 사상과 속성으로부터 실시되어야 한다. 그리고 품질측정값은 공간적 자동상관성을 지니고 있어 데이터의 생산자는 표본설계(sampling scheme)에 있어서 공간자동상관의 존재 및 특성을 정의할 수 있는 표본조사(sampling survey)를 실시하여야 한다.

이와 같은 배경안에서 ISO표준에 따르는 단순임의추출법이 지형공간데이터의 조사에 적용될 수 있다. 공간 자동상관이 충분히 넓은 범위에 있어서 영에 가까운 양(+의 값을 가지면 품질측정값의 공간자동 상관을 무시하기 때문에 단순임의추출법에 의한 모집단의 평균추정값의 분산은 점점 커질 수 있는 것이 지형공간데이터의 품질측정의 매우 일반적인 경향이다. 따라서 이와 같은 추정공간자동상관을 기초로한 표본추출과정은 지형공간데이터를 다년간 다룬 경험이 있는 통계학자의 협조를 받아 설계되어야 할 것이다.

지형공간데이터의 조사는 Fig 2와 같은 표본추출과정을 통해서 실시할 수 있다.

우선 평가받을 데이터(데이터, 부데이터, 사상type, 사상)의 제출을 받아서 데이터를 유사성에 따라 분류하고, 데이터생성단위(사상과 속성)당 불일치발생빈도(likelihood of non-conformance occurrence)의 형태로 데이터를 성질이 비슷한(homogeneous)주제별로 분류한다. 이때 homogeneity는 본래데이터는 거의 같은 품질을 가지며, 데이터생성시스템도 근본적으로 같고, 사상의 복잡성이나 밀도와 같은 불일치발생빈도에 영향을 주는 요소가 근본적으로 같다는 조건하에서 평가하여야 한다.

그리고 이와 같이 분류된 데이터를 다시 여러 로트(lot)로 분류한다. 로트는 item group을 포함한 추상적인 세계안에 있는 공간으로서 정의된다.19) 로트의 크기는 조사단위의 수를 나타내며, 조사단위는 데이터내

Table 2. Results of quality evaluation

	Quality Element			-	Note
	Sub Element			-	-
	Measures			-	-
	Metrics			-	-
Data Set	$\frac{C \text{ or } N}{(\dots)}$			-	-
Feature type	$\frac{C \text{ or } N}{(\dots)}$			-	to be repeated for every relevant feature type
Feature	$\frac{C \text{ or } N}{(\dots)}$			-	to be repeated for every relevant feature type
Feature relation	$\frac{C \text{ or } N}{(\dots)}$			-	to be repeated for every relevant feature type
Attribute type	$\frac{C \text{ or } N}{(\dots)}$			-	to be repeated for every relevant feature type
Attribute	$\frac{C \text{ or } N}{(\dots)}$			-	to be repeated for every relevant feature type
	Det. of * Metrics or reason of not applying			-	-

C : Conformance
 N : Non-conformance
 () : Quality level values in terms of percent non-conforming such as AQL and LQ
 * : Brier Description if newly defined, or not applied

에서가 아니라 추상적 세계의 안에 있는 생산단위와 같은 수가 포함되어 있다고 가정되는 공간단위(spatial unit)를 말한다. 이와 같은 조사단위는 표본추출에 의한 조사단위에 있어서 선택할 수 있는 가장 작은 공간단위이다. 그러나 항목의 실제적인 수를 셀 수 없는 대신에 데이터안의 항목수는 조사단위를 유도하는데 대체할 수 있는 것으로 생산단위의 공간분포추정에는 표본추출방법이 이용될 수 있다. 로트의 공간적 범위는 책임있는 기관에 의해 지정되어야 하며 표본추출의 다른 변수나 로트의 범위지정에 있어서 로트의 크기가 항상 생산과정과 밀접하게 연결되어 있기 때문에 생산과정을 반드시 고려하여 지정하여야 한다.

다시 분류된 로트를 표본추출단위로 분류한다. 이때 표본추출단위는 조사단위에 속해 있는 모든항목을 조사하기 위한 최소공간단위이며 데이터내의 항목수와 각 표본추출수는 거의 같아야 한다.

조사를 위해 임의추출법에 의한 표본추출단위를 선택하고, 선택된 표본추출단위에 속한 항목의 총수는 관계되는 ISO규정에 지정된 표본추출의 수와 같아야 하며 선택된 표본추출단위에 속한 모든 항목을 조사한다.

만일 불일치단위의 수가 AQL이나 LQ (Limiting Quality)에 의해 결정된 지정된 수 보다 클 경우와 측정값의 평균이나 분산값이 AQL에 의해 지정된 한계

조건을 만족하지 않을 경우에는 그 로트는 채택하지 않는다.

마지막으로 표본추출에 의한 조사결과를 품질평가보고서로서 보고한다.

이와 같은 AQL을 이용한 품질을 평가하기 위한 예로서 1994년 핀란드의 LAJA Working group은 지형공간데이터에 관한 등급을 매우우수, 우수, 만족, 부족의 4등급으로 나누었고, 품질요소별로 위치정확도는 0.5, 1, 2, 3, 5, 10, 12.5, 15, 20, 25, 30, 40, 80m으로 14등급 분류하였다. 속성정확도 및 완전성은 전체데이

터내에 부적합한 항목수나 omission의 퍼센트를 표본조사한 결과로부터 AQL을 15로하여 1%, 4%, 15%로 분류하는 방법으로 품질을 평가하였다.

따라서 우리 나라의 지형공간데이터의 교환표준내에도 지형공간데이터의 품질을 정확히 평가하기 위한 지표를 세우기 위해서는 다양한 표본추출방법에 관한 연구와 이 표본조사를 다양하고 광범위하게 실시하여 품질에 관한 자료를 축적하므로써 우리 나라의 사정에 적합한 품질허용기준(AQL)을 정립할 필요성이 있다고 사료된다

Table 3. Quality evaluation methods

	Quality Element			-	Note	
	Sub Element			-	-	
	Measures			-	-	
	Metrics			-	-	
Data Set	Evaluation method applied			-	-	
Feature type	Evaluation method applied			-	-	to be repeated for every relevant feature type
Feature	Evaluation method applied			-	-	to be repeated for every relevant feature type
Feature relation				-	-	to be repeated for every relevant feature type
Attribute type				-	-	to be repeated for every relevant feature type
Attribute				-	-	to be repeated for every relevant feature type
				-	-	

Table 4. Quality requirements in determining conformity of items

	Quality Element			-	-	Note
	Sub Element			-	-	-
	Measures			-	-	-
	Metrics					
Data Set	Quality requirement values, etc			-	-	-
Feature type				-	-	to be repeated for every relevant feature type
Feature				-	-	to be repeated for every relevant feature type
Feature relation				-	-	to be repeated for every relevant feature type
Attribute type				-	-	to be repeated for every relevant feature type
Attribute				-	-	to be repeated for every relevant feature type
				-	-	

3.2.2 간접적인 방법

간접적 평가방법은 품질과 관계되는 외적인 정보로부터 연역법에 의해 추정되는 측정값이나 데이터의 적합도(conformity)에 의한 방법이다. 즉 본래데이터의 품질에 관한 데이터와 생산과정에 있어서의 오차전파 모델 등의 자료로부터 추정하는 방법이다.

연역법에 이용될 수 있는 정보의 예로서 생산에 사용된 본래 데이터의 품질정보, 생산방법등이다. ISO9000 하에서 만들어진 품질에 관한 기록은 데이터 품질측정의 연역에 매우 유용한 입력정보가 될 것이다. 비록 연역법을 통한 품질평가가 직접적으로 품질을 측정하지 않고 어떤 표준과정을 따르지 않지만 표본추출에 의한 조사를 행하는 데 있어서 시간과 비용에 제약을 받을 경우에는 적용가능하다.

연역법의 과정에 있어서 이용된 정보와 참고가 된 정보는 평가법의 신뢰도를 나타내기 위해 품질평가서에 기록되어야 한다.

3.2.3 평가종합

데이터의 전반적인 품질평가는 사상과 속성 등의 요소에 대한 평가결과를 종합하므로써 평가할 수 있다. 평가종합과정에 있어서 구성요소에 대한 significance of non-conformance도 고려하여야 한다. 그리고 평가종합과정도 품질평가보고서에 나타내야 한다.

3.3 데이터 품질 측정의 평가결과보고

평가결과는 일관성있고 표준화된 형식으로된 평가서에 기록 보고하여야 하며 이와 같은 보고형식은 다음과 같다.

우선, 품질평가의 보고를 위한 형식으로서 일치, 불일치 및 AQL 또는 LQ값 등의 품질평가의 결과를 나타내는 보고형식이며 다음 Table 2와 같다. 만일 표준형식이 아닐 경우에는 그 이유를 간단히 기술하고, 또 새로운 데이터품질요소를 추가할 경우는 그 정의를

규정하여 첨가하여야 한다.

다음으로 품질평가과정을 나타내기 위한 형식으로서 평가방법의 보고형식과 항목의 결정을 위한 보고형식이며 평가방법의 보고형식은 만일 표본추출에 근거한 평가방법의 경우 표본설계와 참고가 될 수 있는 정보를 다음 Table 3과 같이 기록하여야 하며 컴퓨터에 의한 자동조사의 경우에는 알고리즘, 제목, 참고정보등이 포함되어야 한다. 그리고 각 항목의 적합성을 결정하는데 필요한 데이터 품질평가는 데이터 생성에 사용된 product spec의 명칭과 함께 Table 4와 같은 형식에 의해서 보고되어야 한다.

4. 결론

본 연구는 기존 및 새로운 지형공간 데이터의 정확성과 신뢰도를 나타낼 수 있는 데이터품질의 정의 및 평가과정을 제시하기 위해 세계 각 국에서 채택하고 있는 데이터 품질요소 정의 및 평가방법을 비교고찰한 것으로서 다음과 같은 결과를 얻었다

첫째, 우리 나라의 국가지형공간데이터 교환포맷의 표준내에 데이터의 품질을 평가하기 위한 요소의 정의는 SDTS의 정의를 그대로 상용하는 것보다는 평가기준을 명확히 제시할 수 있고, 세계적인 표준이 될 ISO/TC211의 안을 적극 수용하는 것이 바람직하다고 판단된다.

둘째, 본 연구를 통하여 보다 합리적인 지형공간정보 데이터의 품질평가과정을 제시하였으며, 이 품질평가과정은 ISO/TC211의 안을 보다 우리 나라 현실에 맞도록 수정한 것으로서 이를 KSDTS에도 적용할 것을 제안하고자 한다. 그리고 데이터 품질측정 결과의 보고형식 등을 더욱더 명확히하여 국립지리원의 수치지도작성규칙등에도 포함시키는 것이 바람직하다고 사료된다.

셋째, 지형공간데이터의 품질을 정확히 평가하기 위한 지표를 세우기 위해서는 다양한 sampling방법에 관한 연구가 요망되며, 이 표본조사를 다양하고 광범위하게 실시하여 품질평가에 관한 자료를 축적하므로써 우리나라의 사정에 적합한 품질허용기준(AQL)을 정립할

필요성이 있다고 사료된다.

넷째, 앞으로의 연구과제는 본 연구에서 제시한 지형공간데이터의 품질요소정의 및 평가과정의 Framework을 바탕으로 하여 대량의 지형공간데이터의 품질을 신속하게 자동평가할 수 있는 컴퓨터 프로그램 작성이 요망된다.

감사의 글

본 연구는 한국과학재단에서 실시하고 있는 해외 Post-Doc.연수의 일환으로 지원을 받아 수행된 연구로서 한국과학재단 관계자에게 감사드리며, 본 연구지도 및 수행에 있어 아낌없는 도움을 주신 현재 ISO/TC211/WG3/NO14의 Leader인, 동경대학생산기술 연구소 R. Shibasaki 교수께 심심한 감사를 드립니다.

참고문헌

1. 村上眞幸, 西城祐輝, "Internetを介した地理情報の所在情報流通とmetadata標準", 日本寫眞測量とRemote Sensing學會誌, Vol.35, No.2, 1996, pp.31~37.
2. Federal Geographic Data Committee(FGDC), <http://fgdc.er.usgs.gov/fgdc.html>.
3. Federal Geographic Data Committee, "Content Standards for Digital Geospatial Metadata", June 8, 1994.
4. Standard Data Transfer Standard(SDTS), "FIPS Publication 173", U.S. Department of Commerce, Technology Administration, National Institute of Standards and Technology, 1992.
5. 한국통신, (주)캐드랜드, "GIS 표준화", KIS GIS 표준화 NO10, 1995
6. 日本 國土地理院, <http://www.gsi-mc.go.jp>
7. (財)日本建設省國土地理院, ISO/TC211 地理情報標準化に関する資料, A1- No.185, 1996.9.
8. CEN/TC287, "Geographic Information Data Description Metadata", 1995.

9. Defence Mapping Agency, "Digital Graphic Information Exchange Standard (DIGEST)", Digital Geographic Information Working Group, 1992.
10. IHO(International Hydrographic Organization), "Guidance on the Treatment of Bathymetric Data Quality on Navigation Chart Products Including the Electronic Navigation Chart", IHO COE Data Quality Working Group, Draft ver2.0, 1995.12.
11. ISO/TC211, Geographic information/Geomatics, NO02, Overview, Ver1.0, 1996.5.
12. 국가 GIS 표준화 분과위원회, GIS 국제 표준화 한국위원회, "국제 GIS 표준 심포지엄", 1995.
13. 한국통신 선로기술연구소, "GIS DB용 데이터 포맷 변환 도구 개발", 1995.12.
14. ISO/TC211/WG1/NO13, Geographic Information, Quality, Ver 1.0, 1996.5.
15. ISO/TC211/WG1/NO14, Geographic Information, Quality Evaluation Procedure, Ver0.2, 1996.9.
16. NJUG(National Joint Utilities Group), Quality Control Procedure for Large Scale Ordinance Survey Maps Digitised OS 1988, Ver 1.0, Pub No.13.
17. ISO2859-0, Sampling procedures for inspection by attributes, Part 0: Introduction to the ISO2589 attribute sampling system.
18. ISO2859-1, Sampling procedures for inspection by attributes, Part 1: Sampling plans indexed by acceptable quality level(AQL) for lot-by-lot inspection.
19. ISO2859-2, Sampling procedures for inspection by attributes, Part 2: Sampling plans indexed by limiting quality(LQ) for isolated lot inspection.
20. ISO3951, Sampling procedures and charts for inspection by variables for percent non-conforming.
21. National Land Survey, "Topographic Data Quality Model", Helsinki, 1995, ISBN 951-48-0133-5.
22. ISO/TC211/WG1/NO15, Geographic Information, Metadata, Ver 1.0, 1996.5.