서울시 공간데이터웨어하우스 품질관리 방안

김윤종, 이석민, 최준영, 김흥열 *
김준기, 김창년, 이창희 **
김학열 ***

* 서울시정개발연구원 도시정보연구센터
** 서울시 지리정보담당관실
*** 서경대학교 도시공학과

1. 연구배경

서울시는 현재 구축되고 있는 GIS 데이터의 내부적 공유체계와 외부적 유통체계의 확립, 데이터의 중복구축 방지 및 활용 등을 위하여 2001년부터 서울시 공간데이터웨어하우스(SDW: Spatial Data Warehouse) 구축 사업을 진행하고 있다. 이에 따라 서울시는 2002년도 이에 대한 시범구축을 수행하였고 2003년부터는 본격적인 데이터웨어하우스 운영 계획을 수립하게 되었다. 2001년도에는 1차로 서울시 공간데이터웨어하우스에서 관리되어야 하는 공통데이터[1]들을 선정하였다. 공통데이터들은 앞으로 서울시 각 업무 부서와 시민들에게 서비스될 예정이며, 이를 위해서는 공간데이터웨어하우스에서 이들에 대한 철저한 품질관리를 실시하여 데이터들의 정확성 및 최신성을 확보하고 서울시 공간데이터웨어하우스의 효과적인 유지관리에 활용하고자 한다. 따라서 본 연구에서는 공간데이터웨어하우스(SDW)의 공통데이터들에 대한 정확성 및 오류 최소화를 위해 데이터 상호간의 정합성(logical consistency) 확보 및 공통데이터의 공유 및 유통을 위한 데이터 최신성 확보를 위한 데이터의 품질관리 방안을 수립하고자 한다.

2. 주요 연구내용

1) 품질관리 범위

2001년도에 서울시에서 선정된 공통데이터간의 일반적 공간관계와 법적 공간관계 등을 바탕으로, 데이터간(레이어별)의 평면적 논리관계(공간관계), 상대적 위치(수평, 수직) 정확성 등을 검사하여 데이터 오류를 확인한 후, 수정함으로써, 데이터간의 상호정합성[2](logical

1) 공통데이터 : 서울시 1/1,000 수치지형도와 현재 각 국·실에서 구축중인 GIS 단위업무시스템들의 데이터를 기반으로 가장 활용도가 높은 65개의 기본 GIS 데이터 레이어들임(예 : 건물, 도로면, 도로중심선, 하수관거, 배수분구, 구경계 등).

2) 정합성 : 데이터 레이어들의 중첩시 지형지물들이 상호간에 공간적으로 상대적 모순이 없이

consistency)을 유지하고 데이터 오류의 확산을 방지하여 데이터 품질을 확보하고자 한다.

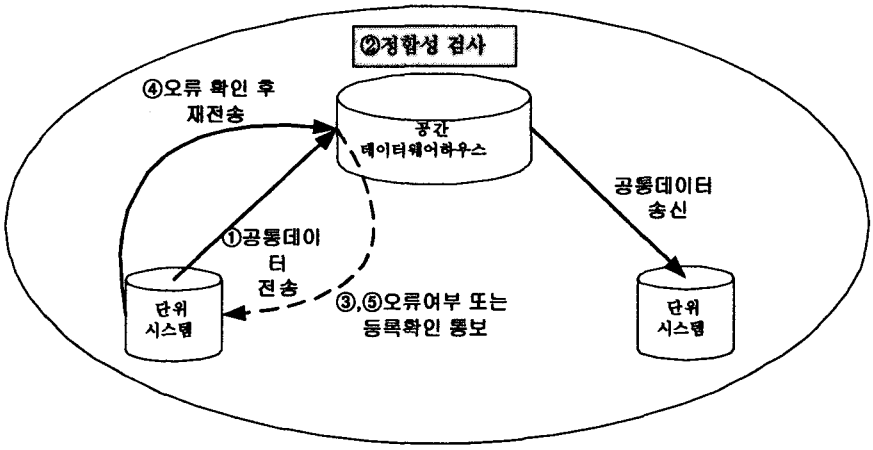
2) 공통데이터의 오류 발생 원인

공통데이터간 정합성 불량으로 발생하는 오류가 제일 중요하며, 데이터 입력오류, 자료갱신 방법과 시기의 차이에서 발생되는 오류, 법적기준과 현실이 상이한 경우 등으로 분류될 수 있다.

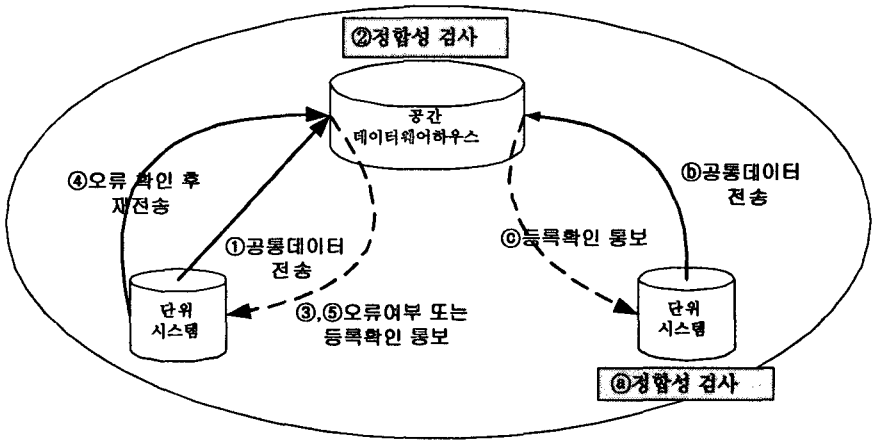
3) 품질관리 절차

본 연구에서는 공통데이터간의 품질관리를 위한 정합성 검사를 검사주체에 따라 공간데이터웨어하우스에서 모든 공통데이터에 대하여 검사를 실시방안과 공간데이터웨어하우스와 단위업무시스템이 검사를 분담하여 실시의 2가지 방안을 제시하였으며, 품질관리 절차는 <그림 1>과 같다.

논리적으로 일치하는 현상(예 : 도로와 건물이 법적인 이격 거리를 확보하고 위치함)



1) 공간데이터웨어하우스에서 정합성 검사를 하는 경우



주) 공간데이터웨어하우스 정합성 절차 : ① →② →③ →④ →⑤
단위업무시스템에서의 정합성 절차 : ⓐ →ⓑ →ⓒ

2) 공간데이터웨어하우스 및 단위업무시스템에서
정합성 검사를 분담하는 경우

그림 1. 정합성 검사 주체별 품질관리 절차

4) 정합성 검사를 위한 공간관계 설정

공통데이터간의 정합성 검사를 위해서는 데이터간 상대적 공간관계가 설정되어야 한다. 따라서 금번 연구에서는 데이터간 일반적 공간관계와 법적 공간관계를 파악한 후, 종합적 공간관계를 설정하는 방법으로 데이터간 공간관계를 설정하였다. 총 90개의 공간관계를 설정하였으며, 최종적으로 시스템 개발 가능 여부 등을 고려하여 실제 정합성 검사가 가능한 68개의 공간관계를 추출하였다(표 2). 이중 건물-도로부속물, 건물-도로(면), 도로부속물-도로시설 등 지상시설물관련 레이어 간의 상대적 공간관계가 30개로 가장 중요하였으며, 지상 및 지하 시설물간의 공간관계는 26개, 지하시설물간의 공간관계(가스-상수도관로 등)는 12개로 나타났다.

표 2. 정합성 검사 가능한 공간관계

경우	구분	공간관계	도형유형	공간관계 설정대상
경우 1 (46)	·지상레이어간의 관계(27)	이격(21)	면대면(6)	·건물 - 도로부속물(면)도로시설(면), 호수·저수지 ·도로부속물(면) - 도로시설(면), 호수·저수지 ·도로시설(면) - 호수·저수지
			면대선(8)	·건물 - 석축상단, 석축하단, 성절토상단, 성절토하단, 옹벽상단, 옹벽하단, 하천중심선 ·도로면 - 하천중심선
			면대점(7)	·호수·저수지 - 맨홀, 빗물펌프장, 소화전, 역사이펀, 오수중계펌프장, 우수토실, 하수맨홀
		포함(6)	면대면(1)	·도로면 - 도로부속물(면)
			면대선(1)	·도로면 - 도로중심선
			면대점(4)	·도로면 - 맨홀, 소화전, 역사이펀, 하수맨홀
	·지상레이어와 지하레이어간의 관계(18)	이격(18)	면대선(18)	·건물-가스, 난방, 상수도관로, 전기, 통신, 하수관거 ·도로면 - 가스, 난방, 상수도관로, 철도, 통신, 하수관거 ·호수·저수지 - 가스, 난방, 상수도관로, 전기, 통신, 하수관거
	·지하레이어간의 관계(1)	인접(1)	선대선(1)	·하수관거 - 하수연결관
경우 2 (20)	·지상레이어간의 관계(3)	이격(3)	면대면(1)	·건물 - 도로면
			면대선(2)	·건물 - 복개하천중심선, 철도
	·지상레이어와 지하레이어간의 관계(8)	이격(8)	면대선(7)	·도로면 - 전기 ·도로시설(면) - 가스, 난방, 상수도관로, 전기, 통신, 하수관거
			선대선(1)	·전기 - 철도
	·지하레이어간의 관계(9)	이격(9)	선대선(9)	·가스 - 전기, 하수관거 ·난방 - 전기 ·상수도관로 - 전기, 통신, 하수관거 ·전기 - 통신, 하수관거 ·통신 - 하수관거
경우 3 (2)	·지하레이어간의 관계(2)	이격(2)	선대선(2)	·가스 - 상수도관로 ·난방 - 상수도관로

주) ( )는 공간관계의 개수임

5) 정합성 검사 실제 적용방안

정합성 검사를 효과적으로 시행하기 위해서는 최종 정합성 검사 결과에 따른 데이터 품질 확보 기준을 우선 설정하여야 한다. 데이터 품질확보는 고품질 확보를 위한 실측(측량 등)실시 방안 및 내부업무 활용에 요구되는 품질확보 방안 등이 있다. 데이터의 고품질 확보를 위한 모

든 지형지물들에 대한 실측은 현실적으로 불가능하므로 서울시에서는 내부업무 활용에 요구되는 품질확보 방안이 현 시점에서 타당한 방안으로 판단된다. 내부업무 활용에 요구되는 품질확보 방안은 첫째, 기준데이터[3]를 선정하여 정합성 검사를 실시하는 방법과 둘째, 기준데이터 없이 검사하는 방법으로 구분될 수 있다. 기준데이터를 선정하여 정합성 검사를 실시하는 방안을 도입하기 위해서는 우선 오류가 가장 적은 기준데이터가 선정되어 이를 기준으로 검사대상 데이터에 대한 오류를 판정하여야 한다. 그러나 현재 GIS 데이터들이 구축 중인 서울시 환경에서는 이러한 기준데이터 선정은 무리가 있으며, 장기적 검토가 필요하다. 따라서 기준데이터 없이 정합성 검사를 실시하는 방안이 현재 서울시의 GIS 구축환경에서는 실행 가능한 것으로 판단된다.

3. 결론

본 연구에서는 서울시 공간데이터웨어하우스 구축과 관련하여 데이터들의 품질관리 절차의 확립하고, 품질관리 수행 주체 결정, 단위업무시스템에서의 공통데이터갱신·검사시기결정·정합성 검사실시와 공간데이터웨어하우스에 데이터 등록·검사시기결정·정합성검사 등 단계별 품질관리 절차를 마련하였다. 또한 각 단위업무시스템에서 수정 갱신되는 데이터간의 정합성 검사를 위한 공간관계 설정하고자 데이터간 일반적 및 법적 공간관계를 파악한 후 이를 종합하여 90개의 공간관계를 설정하였으며, 최종적으로 시스템 개발 가능 여부 등을 고려하여 실제 정합성 검사가 가능한 68개의 공간관계를 추출하였다(총 대상 29개 레이어).

데이터 품질확보 기준 설정은 서울시에서 현재 공간데이터가 구축중인 점을 감안하여 품질확보 기준은 우선 내부업무 활용에 요구되는 품질확보를 기준으로 하며, 소유권 분쟁, 보상 등의 민원업무처리는 좀더 고도의 품질확보 기준(측량 실시)을 마련하는 것이 필요하다. 품질확보 방법은 현재 기준데이터 없이 정합성 검사를 실시하여 검사결과를 각각의 검사대상데이터를 생성한 단위업무시스템에 통보함을 원칙으로 하나, 그러나 장기적으로는 기준데이터를 선정하여 이를 기준으로 공통데이터간 정합성 확보 및 데이터 수정을 실시하는 것이 필요하다.

3) 기준데이터 : 정합성 오류검사를 위해 기준으로 설정되는 데이터임. 또한 단위업무시스템에서 관련 데이터 입력을 위해 기준이 되는 데이터로도 사용될 수 있음.