

DEVELOPING ENTERPRISE GIS-BASED DATA REPOSITORIES FOR MUNICIPAL INFRASTRUCTURE ASSET MANAGEMENT

Mahmoud Halfawy¹ and Roberto Figueroa²

ABSTRACT

Efficient integration and management of assets life-cycle data is the key to successful management of municipal infrastructure systems. Lack of data integration can result in significant inefficiencies and suboptimal operational and long-term asset management strategies. Integrated asset management systems can help municipalities overcome these inefficiencies and improve the coordination and cost-effectiveness of asset management decisions. Challenges of data integration and management are by far the biggest impediment to implementing successful asset management strategies.

This paper describes an approach for the development of centralized enterprise-wide shared data repositories to integrate asset spatial and non-spatial data. The data repository would serve as a medium for sharing and integrating data across all municipal departments and software systems. An implementation case study is also presented. The case study demonstrates the implementation of a centralized asset data repository in the City of Regina, using Oracle DBMS and ESRI's ArcSDE gateway. The process of schema mapping and data conversion, and the transition from using paper records, to CAD maps, to departmental GIS, and finally to multi-user enterprise GIS-based asset data repository is described.

KEY WORDS

Municipal infrastructure, asset management, data integration, municipal GIS, asset life-cycle data.

¹ Research Officer, Centre for Sustainable Infrastructure Research, National Research Council, 6 Research Drive, Regina, SK, Canada, S4S 7J7, Phone +1 306/780-5396, FAX +1 306/780-3421, Mahmoud.Halfawy@nrc.ca

² GIS Coordinator, The City of Regina, 2476 Victoria Ave., Regina, SK, Canada S4P 3C8, Phone +1 306/777-7575, FAX +1 306/777-6806, RFiguroa@regina.ca

INTRODUCTION

Management of municipal infrastructure assets involves many complex, inter-dependent, and data-intensive processes. Successful implementation of asset management strategies largely depends on: (1) the efficiency to share, access, and manage the asset life-cycle data; and (2) the ability to efficiently support and coordinate the multi-disciplinary work processes at the operational and strategic levels. Lack of data integration in managing municipal infrastructure assets can result in significant inefficiencies and suboptimal operational and long-term asset management strategies. Implementation of integrated asset management systems can significantly help municipalities overcome these inefficiencies and improve the coordination and cost-effectiveness of asset management decisions (Halfawy 2004).

Municipal asset data are typically characterized by their sheer size, complexity, inter-dependencies, and dynamic nature. These data typically exist in disparate formats across multiple data sources and software systems. Integrating these data into a consistent and unified form is recognized as a critical step towards successful asset management. However, data integration has proved to be one of the major challenges municipalities currently face.

During the past two decades, a plethora of municipal asset management software has been in use. However, the majority of these tools were developed to function as stand-alone systems with limited or no capability to share and exchange information with other systems. The proliferation of these tools has created the familiar “islands of information” phenomenon in most of our municipalities, which can be primarily attributed to the use of inconsistent data models across these disparate software tools.

Asset management decisions typically require the integration of data of different formats from multiple disparate and distributed sources and software systems. Efficient data integration and management can significantly improve operational efficiency and cost-effective asset management decision-making at the operational, tactical, and strategic levels. Data integration can be defined as “the process of combining or linking two or more data sets from different sources to facilitate data sharing, promote effective data gathering and analysis, and support overall information management activities in an organization”(FHWA 2001). The FHWA report also identified a number of benefits for data integration and sharing, which included: data availability/accessibility; timeliness; accuracy, correctness, and integrity; consistency and clarity; completeness; reduced duplication; faster processing and turnaround time; lower data acquisition and storage cost; informed and defensible decisions; and integrated decision-making.

Asset spatial data constitute the core of most municipal information systems, and are central to many asset management decision-making processes. Infrastructure asset data are typically identified, associated with, or referenced by their geographic locations and spatial relationships. As a result, there has been an increasing realization of the useful role that Geographic Information Systems (GIS) and spatial data analysis can play to support asset management processes (Vanier 2000, Halfawy 2004). A recent review of commercial municipal asset management systems (Halfawy et al. 2005) has found that GIS is becoming an integral component in almost all modern municipal information systems, and that the vast majority of existing software tools has already implemented GIS functionality either internally or by supporting interfaces with commercial GIS systems.

Most municipal GIS implementations to date have primarily followed a departmental or a service group approach (GITA 2005). In both cases, spatial data have typically been stored and managed in personal or departmental databases that limited the sharing and editing of the data to one user group. The increasing demands to share spatial data for various municipal applications has created a strong need to develop enterprise-wide GIS. Recent availability of low cost server hardware and software has made the development of such systems more feasible and cost-effective. Recent reports indicate that there is an increasing trend across various industries to deploy enterprise GIS systems (GITA 2005, IDC 2005). This paper discusses an approach to leverage the role of enterprise GIS to better support the development of integrated infrastructure asset management systems and strategies.

ROLE OF INTEGRATED ASSET DATA REPOSITORIES IN MUNICIPAL INFRASTRUCTURE MANAGEMENT

The proposed approach involves the development of integrated and shared GIS-based asset data repositories based on integrated data models that represent the spatial structure of the infrastructure asset, and also incorporate (or cross-reference) various aspects of asset life-cycle information, such as inspection, performance, rehabilitation, and cost data (Figure 1). The repository can potentially improve the efficiency, cost-effectiveness, and coordination of various asset management processes.

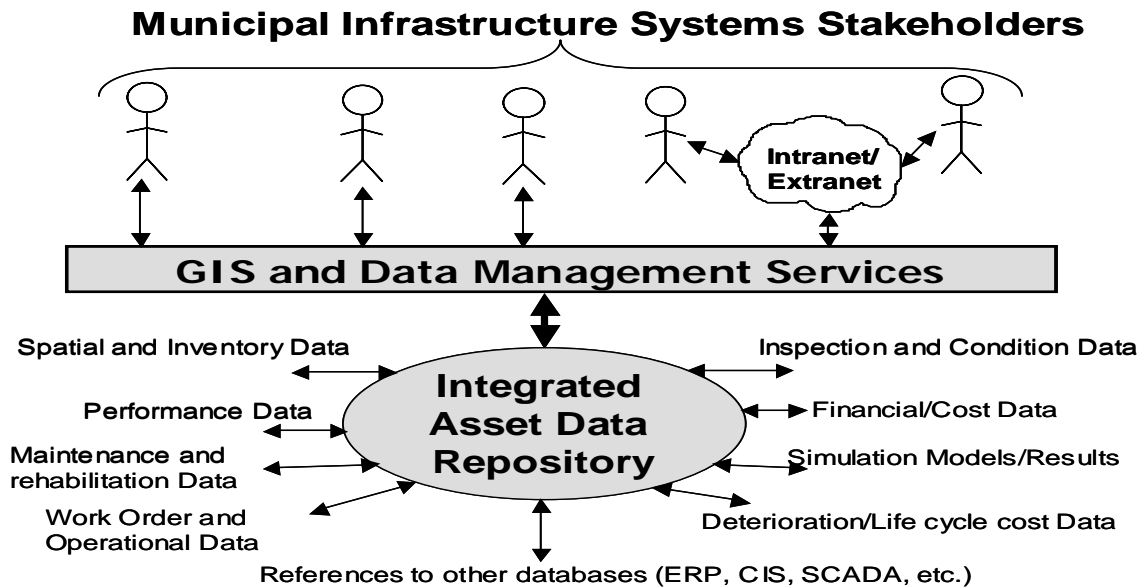


Figure 1: Role of data repositories as integrators of asset data and processes

The integrated GIS-based asset data repository is built on top of a relational Database Management System (DBMS), and therefore can provide a wide range of data sharing, integration, and management services, such as version management, multi-user concurrent access and edit, security and authorization, and metadata services. The repository would also ensure the consistency and integrity of asset data, and serve to bind various perspectives of

the data into a comprehensive and consistent view of the entire infrastructure system. By providing a single point of access to all asset data, the repository can significantly improve the collection, organization, management, and distribution of data throughout the asset life-cycle. The GIS interface to the repository will further enhance users' ability to explore, query, and analyze asset data in a spatial context.

The repository can support the management of inter-dependent infrastructure systems, such as water main, sewer, and road networks in an integrated manner. These different infrastructure networks are typically managed separately, and the data are stored and managed in separate, and possibly incompatible, databases. The integrated repository can maintain cross-references and relationships between different infrastructure systems, which can support the investigation of the interaction between different co-located or overlapping assets. Various work groups will be able to come together with a consistent view of the data, in an integrative and multidisciplinary asset management approach.

The repository enables the interoperability and efficient data sharing between disparate legacy asset management systems (e.g., condition assessment, renewal planning systems) and other enterprise-wide systems (e.g., Enterprise Resource Planning, ERP, systems, financial systems). The repository can also support the use of distributed data sources, and enables access to these sources locally or over the Internet/Intranet (in a client-server or web-based architecture) by multiple distributed client applications within or across the organizational boundaries. Some other benefits of centralized asset data repositories may include:

- Enabling efficient storage, correlation, indexing, query, and analysis of asset data, and enable concurrent retrieval and editing of these data by multiple applications;
- Supporting the coordination and streamlining of asset management processes, which can lead to increasing the operational efficiency and enhancing the communication among different municipal departments and stakeholders.
- Enabling data reusability and sharing, and thus help eliminate the duplication of efforts, and potential inconsistency and redundancy, in gathering, validating, and storing asset data.
- Promoting the use of consistent, integrated, and standardized vendor-neutral data models and formats;
- Enabling easy integration of legacy software tools, previously isolated in separate silos, into a unified and coherent enterprise-wide environment (Halfawy et al. 2002).

DATA MODELS FOR DEFINING THE ASSET DATA REPOSITORY SCHEMA

Currently, municipal infrastructure data are mostly shared and exchanged between different software tools in two main ways. First, through the use of translators to map the data between different vendor-specific data models and formats. The translation process typically involves a tremendous amount of redundant data retrieval, interpretation, and re-entry, and is known to be prone to mapping and interpretation errors. Moreover, the tool-specific nature of the translators imposes unnecessary constraints by requiring the use of proprietary vendor-specific data models and software systems. Second, through accessing a set of application

programming interfaces (API) to access the application's internal data model and to input or extract data directly to and from the application. Although some of these APIs appear to be de-facto industry standard, the majority of the APIs are proprietary and tool-specific. The use of proprietary tool-specific methods is a clear obstacle for asset data integration. Using centralized asset data repositories based on vendor-neutral standard data models seems to be the most viable option for asset management data integration and software interoperability.

A major challenge in building centralized data repository is the need to develop a data model and a corresponding database schema to represent and integrate asset life-cycle data in a unified, comprehensive, and preferably standardized, manner. Software tools that comply with the standard can be easily integrated into the repository without the need to develop tool-specific adapters to translate asset data to and from the repository data model, and will thus facilitate the development and deployment of integrated asset management systems.

During the past decade, there has been a surge of activities to develop standard integrated data models. A number of these efforts have produced a set of widely accepted spatial data standards. Most notable are the work of the OpenGeoSpatial Consortium, OGC (www.opengeospatial.org) and the ISO/TC 211 (www.isotc211.org). Parallel to these efforts, a number of initiatives were launched to standardize municipal infrastructure data. Most notable is the work of the Federal Geographic Data Committee, FGDC (www.fgdc.gov), which was later merged into the CADD/GIS Technology Center (tsc.wes.army.mil). This work produced a number of data standards such as the Spatial Data Content standards, the Spatial Data Transfer Standards (SDTS), the metadata standards, and the Spatial Data Standard for Facilities, Infrastructure, and Environmental Applications (SDSFIE). Other related efforts included industry-led initiatives (e.g. the LandXML standard, www.landxml.org), public work organizations initiatives (e.g. the Municipal Data Standard, MIDS, www.tricom.org), and software vendors initiatives (e.g. ESRI utilities data models). An overview of these data standards was provided in (Halfawy 2004).

Many of the aforementioned data standards have been in use for several years, and thus have reached a high level of maturity. However, most of these data models have primarily focused on the spatial and physical aspects of the infrastructure assets, and lacked other critical life-cycle aspects such as condition, performance, deterioration patterns, maintenance and rehabilitation, and cost. Moreover, many of these data models have not defined implementation or format standards for data exchange.

Recent developments of XML standards and technologies potentially provide new opportunities to develop integrated data models, as well as efficient data encoding and exchange mechanisms. Work is currently underway at the NRC's Center for Sustainable Infrastructure Research (CSIR) to develop integrated data models for water and wastewater infrastructure systems compliant with the new OGC Geographic Markup Language (GML) 3.1 schema. The GML schema is planned to be released as ISO 19136 standard. The data models use the GML elements and data types to encode the spatial asset data, and define new elements and data types for other life-cycle non-spatial data. The integrated data models are intended to represent a comprehensive view of the infrastructure system and to act as the glue that binds together various perspectives of the infrastructure information. In addition to representing the asset spatial and structural characterizes, the data models also represent a wide range of asset life-cycle data (e.g. condition, rehabilitation data, cost).

EXAMPLE IMPLEMENTATION: THE EXPERIENCE OF THE CITY OF REGINA

Implementation of an enterprise GIS-based asset management system is a complex and long process that typically proceeds in multiple phases, and lasts for several years. The time and resources needed typically depend on the size of the municipality, available resources, and the level of data availability and reliability. In early 1990s, the City of Regina started its GIS implementation effort by the development of the Land-Based Information System (LBIS), where all City spatial data and related processes were identified and analyzed. The LBIS development included tasks such as the conversion of maps and data sets into digital format, setting a standard for street address definition, and ownership parcel layer maintenance. The implementation has evolved towards the goal of deploying a fully integrated data repository to support various infrastructure management processes. The following sections describe the implementation process evolution starting from the use of paper records to the development of enterprise GIS-based asset data repository.

DATA CONVERSION: FROM PAPER RECORDS TO CAD MAPS

The growing needs to produce infrastructure records (estimated at 50,000 plans/year) for various departments revealed several inefficiencies in the use of paper-based records. The City launched a five-year initiative to convert infrastructure records into digital format. The Infrastructure Record Conversion (IRC) project started by re-drawing the infrastructure record plans (about 3,500 sheets) and saving the data in CAD format. Instead of scanning the drawings, the project team decided to redraw all plans to enable better quality assurance and control. As the demand for CAD data grew higher, CAD records exhibited some serious limitations. First, data access was limited to CAD users, while others had to rely on mass printouts kept at the drafting counter. Second, the CAD models represented features as geometric primitives that lack the semantics needed to verify and validate the data. Third, most CAD systems exhibit limited data management functionality, which impeded the ability to efficiently maintain the linkages between the spatial objects and their attribute data.

MIGRATING CAD MAPS TO GIS

Unlike traditional CAD models, GIS models emphasize the definition and use of semantic rich object-based models that are linked to Relational DBMS to store and manage the attribute data. By integrating spatial and non-spatial data, GIS models can enable efficient and automated data verification and validation; both are critical for ensuring the quality and reliability of the asset data. Moreover, the client-server architecture of most GIS software enables browser-based thin clients to efficiently access the data over Intranet/Extranet networks. The browser-based architecture can support the mass deployment of spatial data across various departments and stakeholders within municipalities in a cost-effective manner.

The project team adopted the ESRI ArcGIS architecture, and started by developing a geodatabase for domestic and storm sewer data. The ESRI geodatabase data model is an object-relational schema that includes relational tables, feature classes, relationship classes, topologies, geometric networks, raster datasets and raster catalogs, and data semantics and behaviour rules. The geodatabases data model enables the implementation of validation rules to enforce integrity constraints on the data, and employs relational DBMS SQL functions to

support data access, update, and management processes. The geodatabase data model allows the definition of “custom objects” that embody user-defined semantics. For example, instead of representing a sewer as a generic line feature, a custom sewer feature can enforce subtypes, domains, and relationship rules to ensure the validity of the feature. The geodatabase model also enables the implementation of more complex spatial relationships such as networks, topologies, and terrains.

The geodatabase schema was defined using the Unified Modeling Language (UML), which was then input to the ArcCatalog’s Schema Wizard to generate the database tables, fields, and data types. As an example of the schema development process, the Domestic Feature dataset was divided into six feature classes – Pipe, Fitting, Manhole, Structure, Liftstation and Pumpstation. Each of these feature classes was further divided into subtypes. Subtypes are integer-valued attribute that are used within a feature class or a table to emulate a class hierarchy. Subtypes inherit all the properties of the feature class. Various validation rules (e.g. domains, relationship rules) can be applied to individual sub-types. Figure 2 shows the definition of the Domestic Sewer Line (DSLLine) feature class. Three subtypes in the DSLLine Feature Class were defined: Main, Trunk and Force. All three subtypes have the same attributes, but are classified differently because of their diameter. Domains (or range of valid numbers or list values) were defined on the fields in the feature class. Two Coded Value Domains were defined: PipeMaterial and MainPipeDiameter. The PipeMaterial domain is attached to the Material field in the Feature Class, while the MainPipeDiameter domain is attached to the Diameter field in the Main subtype (Weston et. al., 2001).

A geometric network was then created to establish connectivity rules and cardinality for each object in the model. These rules establish which features can be connected, and the restrictions of the number of edges connecting to a junction, or the number of junctions connecting to an edge. Some examples of the types of rules and cardinality applied were: a manhole must connect to at least two mains, a coupling can only have two pipes connected to it, and an end cap can only have one pipe connected to it.

After the geodatabase schema was designed, domestic and storm sewer data were converted from the AutoCAD drawings into the geodatabase schema. The objects (or features) and attributes were already established in AutoCAD with attributes, attribute domains, layers and blocks (or symbols). Data were exported from AutoCAD to ESRI shape files, and then imported to the geodatabase using ESRI’s ArcCatalog’s Simple Data Loader. After the data were loaded, a validation was performed to identify the number of existing errors such as not falling into a feature’s applied domain, or the violation of network connectivity and cardinality rules. Network analysis tools were also used to determine if there were ‘holes’ or disconnected areas in the network topology.

After migrating the domestic and storm sewer data into the geodatabase and correcting the errors flagged in the validation process, the data maintenance of these layers (changes, additions, or deletions) were mainly performed in ESRI’s ArcMap. Junction subtypes (manhole, fitting, etc) and attribute information were also added and edited in ArcMap. At a later stage, the water distribution system and pavement layers were added to the geodatabase. In addition to infrastructure layers, the geodatabase also included other corporate layers such as parcels, land use, digital elevation maps, and aerial ortho-corrected photos.

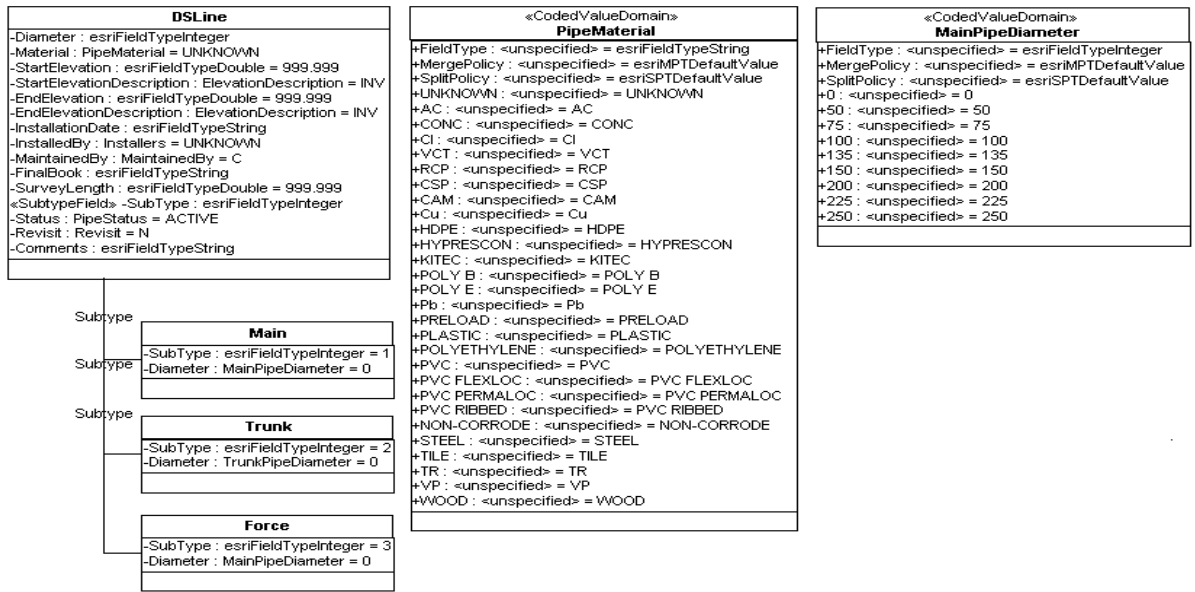


Figure 2: UML Domestic Sewer Line Feature Class Definition (Weston et al. 2001)

Up to this phase, the project team has dedicated most of its effort to address data collection, conversion, correctness, and validation challenges. The geodatabase implementation was in the form of a personal geodatabase (or PGDB). PGDB is stored in a Microsoft Access (MDB) file, and can only be edited by one user at a time. Moreover, the size of PGDB is limited to 2 GB, which limits its scalability. These limitations, combined with the need to leverage the use of spatial data across various departments and processes, made the project team move to migrate the GIS data to an enterprise centralized system.

MIGRATION TO ENTERPRISE GIS AND ASSET DATA REPOSITORY

In 2000, the project team started to build and maintain an enterprise GIS system that can support the creation of multi-user versioned geodatabases. The enterprise GIS implementation was based on ArcSDE and Oracle relational DBMS. The enterprise implementation was also intended to support mass and field deployment of up-to-date spatial data using ArcIMS technology (Figuroa and Stusek, 2001).

ArcSDE serves as a gateway that connects distributed client applications to a relational DBMS (e.g. Oracle, Microsoft SQL Server, IBM DB2, and Informix) (ESRI 2004). ArcSDE implements the ISO 19107 Spatial Schema standard, which was originally defined by the OpenGeoSpatial Consortium. ArcSDE also implements the ISO 19125 standard for “Simple Feature Access,” which defines SQL data types and functions for representing, exchanging, and managing spatial data in relational databases.

ArcSDE creates and manages a set of tables (or data dictionary). These tables store metadata about spatial data such as spatial references, feature class names and structures, and spatial indexing (ESRI 2004). Mapping of the geodatabase data model to and from client-specific data models is achieved by using adapters (Figure 3). ArcSDE can store feature geometric data in a binary field in a relational table in one of two formats: the ArcSDE compressed binary format (in Oracle and SQL Server), or the native DBMS spatial column

types (in Oracle spatial, Informix, and IBM DB2). ArcSDE translates data between the relational representation and the object-based geodatabase representation.

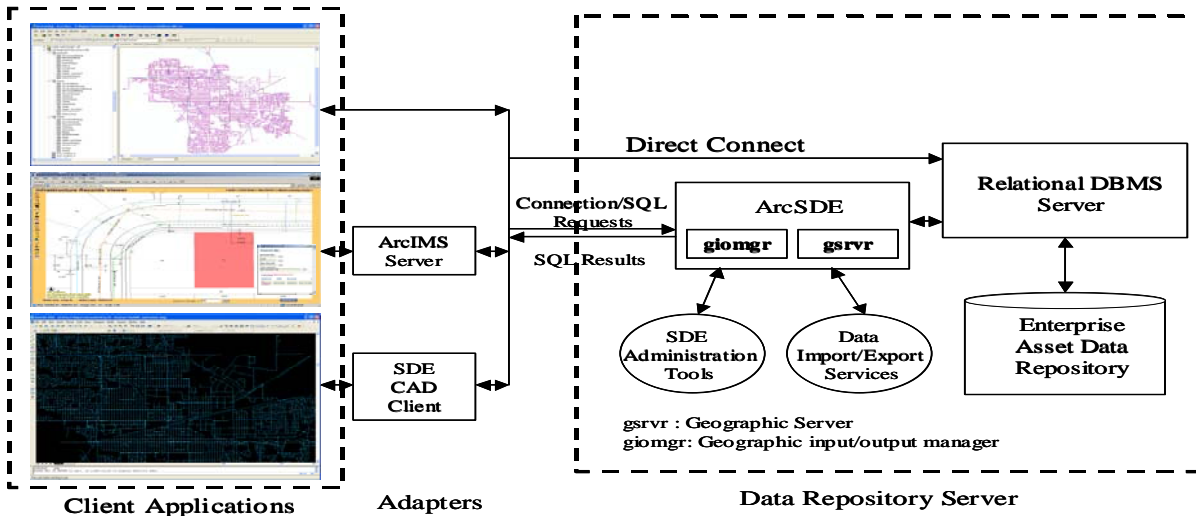


Figure 3: Architecture of the Enterprise GIS-Based Infrastructure Asset Data Repository

Client applications communicate with ArcSDE (over a Transmission Control Protocol/Internet Protocol, TCP/IP, connection) by passing SQL statements to retrieve, store, or update the data in the DBMS. The DBMS returns a set of rows to the client. To increase the efficiency of data retrieval, clients need to provide spatial or attribute filters. Attribute filters are defined in the form of WHERE clauses of the SQL statements. The WHERE clause can join the ArcSDE business table with other related tables to retrieve features based on conditions defined in these joined tables. Spatial filters are used to retrieve data based on spatial condition (i.e. location). ArcSDE or the DBMS constructs and applies the spatial filter (in the form of a WHERE clause) to reduce the number of retrieved features. The DBMS uses spatial indexing to reduce the search and retrieval time of spatial data. Two commonly used methods for spatial indexing are supported: the quadtree method (used by ArcSDE, Oracle spatial, and IBM DB2), and the R-tree method (used by Informix and Oracle spatial).

Multiple users can concurrently connect to, query, edit and update asset data. Appropriate permissions can be granted to different users to satisfy their varying data access requirements. Read-only (or SELECT permissions) are granted to users who only need to access the data; while INSERT, DELETE, and/or UPDATE permissions are granted to users who make changes to the data.

The repository supports maintaining multiple versions of the database. This would enable keeping track of data update history, and rolling back changes if need arises. To optimize the use of storage resources, changes are stored in delta-tables, and different versions are maintained using these tables instead of replicating the entire database. Multiple versions can be later reconciled by integrating changes to one version. If all version changes are consistent, the reconciliation process is easily achieved by applying the changes stored in delta-tables. However, if the changes are inconsistent (e.g. different values of a feature attribute), a conflict would rise and the user will be prompted to choose which value to store.

CONCLUSIONS

Integrated GIS-based asset data repositories can significantly improve the availability and consistency of the asset data across different software systems, integrate data across various disciplines, and facilitate the flow and exchange of asset information. This can potentially result in reducing or eliminating inefficiencies of information access and exchange, and thus lead to cost-effective and more efficient operational and strategic decisions. This paper described an approach for implementing centralized GIS-based asset data repositories, and presented specific primitives (e.g. data models, software architecture) required to realize that implementation.

Future work will need to focus on the development of standard integrated data models. These standards should build on previous efforts, be based on industry wide consensus, and be defined through consortia of public works professionals, software developers, and other interested stakeholders. The deployment of asset data repositories, with their advanced data integration and management capabilities, will be a key enabler to develop and adopt more efficient and cost-effective asset management systems.

ACKNOWLEDGMENTS

This work is part of a collaborative initiative between the National Research Council and the City of Regina. The authors would like to thank David Hubble, Yehuda Kleiner, and Darryl Dormuth for their valuable comments.

REFERENCES

- FHWA. (2001). "Data Integration Primer," Federal Highway Administration, available online at <isddc.dot.gov/OLPFiles/FHWA/010393.pdf>.
- GITA. (2005). "2005 Geospatial Technology Report," Geospatial Information & Technology Association.
- IDC. (2005). "ESRI: Extending GIS to Enterprise Applications," available online at <www.esri.com/library/whitepapers/pdfs/idc_enterprise_apps_feb_2005.pdf>.
- ESRI. (2004). *Understanding ArcSDE*, ArcGIS product documentation.
- Halfawy, M.R., David Pyzoha, and Taymour El-Hosseiny, (2002). "An Integrated Framework for GIS-Based Civil Infrastructure Management Systems," *Proc., CSCE 2002*, the Canadian Society for Civil Engineers, Montreal, Canada.
- Halfawy, M., (2004). "The Interoperability of Geographic Information Systems for Municipal Asset Management Applications", Municipal Infrastructure Investment Planning (MIIP) Project Report, Part II, National Research Council Canada.
- Halfawy, M, Newton, L, and Vanier, D. (2005). "Municipal Infrastructure Asset Management Systems: State-of-the-art Review", *Proc., CIB W78*, Dresden, Germany.
- Vanier, D., (2000). Municipal Infrastructure Investment Planning (MIIP) Project: Statement of Work, Inst. for Research in Construction, National Research Council Canada.
- Figuroa, R., J. Stusek (2001) "Regina Enterprise Geographic Information System (REGIS)", *Proc., GeoSASK 2001*, Regina, SK, Canada.
- Weston, A., G. Henry, D. Poshen (2001). "Geodatabase Modelling for Infrastructure Records at the City of Regina," *Proc., GeoSASK 2001*, Regina, SK, Canada.