Standard data models for interoperability of municipal infrastructure asset management systems[1]

Mahmoud R. Halfawy, Dana J. Vanier, and Thomas M. Froese

Abstract: Efficient management of infrastructure assets depends largely on the ability to efficiently share, exchange, and manage asset life-cycle information. Although software tools are used to support almost every asset management process in municipalities, data exchange is mainly performed using paper-based or neutral file formats based on ad hoc proprietary data models. Interoperability of various asset management systems is crucial to support better management of infrastructure data and to improve the information flow between various work processes. Standard data models can be used to significantly improve the availability and consistency of asset data across different software systems, to integrate data across various disciplines, and to exchange information between various stakeholders. This paper surveys a number of data standards that might be used in implementing interoperable and integrated infrastructure asset management systems. The main requirements for standard data models are outlined, and the importance of interoperability from an asset management perspective is highlighted. The role that spatial data and geographic information systems (GIS) can play in enhancing the efficiency of managing asset life-cycle data is also discussed. An ongoing effort to develop a standard data model for sewer systems is presented, and an example implementation of interoperable GIS and hydraulic modeling software is discussed.

Key words: data standards, municipal infrastructure, asset management, data models, interoperability.

Résumé : La gestion efficace des infrastructures dépend en grande partie de l'habileté à partager, à échanger et à gérer efficacement l'information sur le cycle de vie des actifs. Bien que des outils informatiques soient utilisés pour supporter le processus de gestion de presque tous les actifs municipaux, l'échange de données se fait surtout par des supports papier ou des formats de dossier basés sur des modèles ponctuels de données exclusives. L'interopérabilité des divers systèmes de gestion des actifs est essentielle à la gestion des données d'infrastructures et les échanges d'information entre les divers processus de travail. Des modèles standards de données peuvent être utilisés pour améliorer de manière significative la disponibilité et la cohérence des données sur les biens dans les différents systèmes logiciels, intégrer les données de divers domaines et échanger des informations sur le cycle de vie des actifs entre les divers intervenants. Cet article étudie un certain nombre de normes de données qui peuvent être utilisées pour implanter les systèmes de gestion des infrastructures interopérables et intégrés. Sont également soulignées les principales exigences pour les modèles standards de données et l'importance de l'interopérabilité du point de vue de la gestion des actifs. Cet article aborde également le rôle que les données spatiales et des systèmes d'information géographique (« GIS ») peuvent jouer pour accroître l'efficacité de la gestion des données du cycle de vie des actifs. Nous présentons les travaux en cours pour développer un modèle standard de données pour les réseaux d'égouts ainsi qu'un exemple d'implantation des logiciels de modélisation « GIS » et hydrauliques interopérables.

Mots clés : normes de données, infrastructure municipale, gestions des actifs, modèles de données, interopérabilité.

[Traduit par la Rédaction]

Received 28 March 2005. Revision accepted 3 October 2005. Published on the NRC Research Press Web site at http://cjce.nrc.ca on 3 March 2007.

M.R. Halfawy[2,3] and T.M. Froese. Department of Civil Engineering, The University of British Columbia, 6250 Applied Science Lane, Vancouver, BC V6T 1Z4, Canada.
D.J. Vanier. Institute for Research in Construction, National Research Council of Canada, 1200 Montreal Road, Ottawa, ON K1A 0R6, Canada; and Department of Civil Engineering, The University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

Written discussion of this article is welcomed and will be received by the Editor until 30 April 2007.

[1]This article is one of a selection of papers published in this Special Issue on Construction Engineering.
[2]Corresponding author (e-mail: Mahmoud.Halfawy@nrc-cnrc.gc.ca).
[3]Present address: Centre for Sustainable Infrastructure Research, Institute for Research in Construction, National Research Council of Canada, 6 Research Drive, Suite 301, Regina, SK S4S 7J7, Canada.

 doi:10.1139/L05-098

1. Introduction

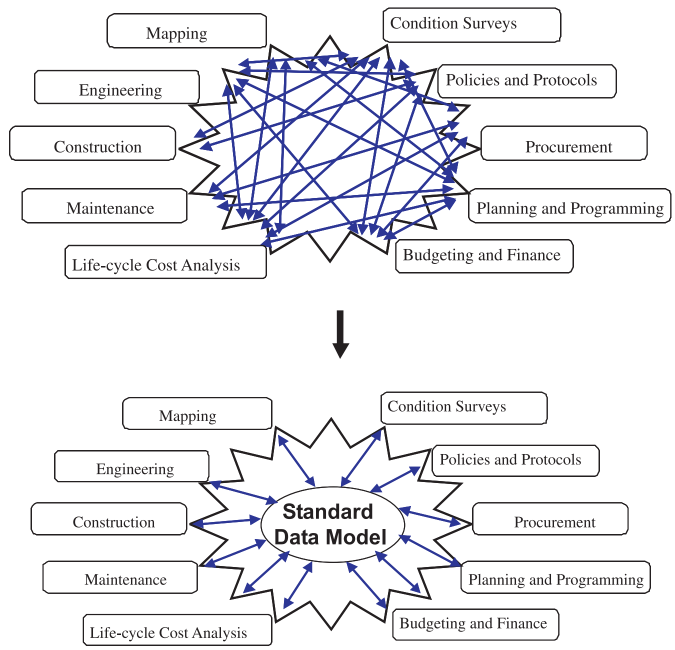
Many municipalities use asset management systems to store and manage information and to support tactical and strategic decisions regarding the operation, maintenance, rehabilitation, and replacement of their infrastructure. Implementation of efficient and cost-effective management strategies largely depends on the ability of these systems to share and exchange asset life cycle information.

Municipal asset data are typically characterized by their long life cycle, sheer size, complexity, inter-dependencies, and dynamic nature. The primary function of a municipal asset management system is to maintain the accuracy, consistency, and integrity of the asset data. Many municipalities have made significant investment in implementing software systems to address the increasing complexity of storing and maintaining the data associated with their infrastructure assets. In the past decade, significant advances have been made in developing management systems for municipal assets, such as sanitary and stormwater sewers, water supply distribution systems, pavements, and bridges (Halfawy et al. 2000, 2002; Vanier 2001). These asset management systems support a wide range of functions, such as collecting inventory and condition data, and supporting maintenance operations. A number of municipalities have also adopted innovative methods for assessing asset condition, analyzing life-cycle costs, planning and prioritizing maintenance operations, simulating and predicting performance, and evaluating alternative technical and economic policies (Halfawy 2004; Vanier and Rahman 2004).

The use of infrastructure asset management applications has significantly improved the operational efficiency and maintainability of the assets. However, the majority of these applications were developed to function as stand-alone systems and, therefore, have limited or no capability to share and exchange information. Each application typically uses a proprietary data model and stores data in a proprietary file format, which, in most cases, cannot be accessed by or shared with other software applications. Users typically have to translate the data from one representation and format to another so the data can be used across different applications. The translation process involves many time-consuming, error-prone, and inefficient activities: getting output from one application in a paper or digital format, reinterpreting the output, and then reentering the data into another application. The lack of interoperability and integration between municipal asset management systems has resulted in problems and inefficiencies in accessing, exchanging, and managing data (Halfawy 2004). Experience shows that the use of proprietary data models and formats has created many obstacles to improving the availability, quality, and reusability of data.

Standardizing the representation of municipal data would allow interoperability and provide users with the capability to reuse existing data, coordinate work processes, and share vital information in an efficient and effective manner (Fig. 1). Standardized data models would help define common and consistent data structures and semantics using vendor- and technology-neutral data encoding and exchange formats. Ideally, a standard data model would also provide an integrated schema for representing and exchanging data across all asset life-cycle phases, including planning, design, construction, operation, and maintenance.

Fig. 1. Role of standard data models in supporting integration and interoperability.



To facilitate the representation, querying, analysis, and management of data, asset management systems should incorporate spatially based asset data repositories that can integrate spatial and nonspatial attributes. A key challenge to developing these repositories is the lack of standardized and integrated representation of asset spatial and nonspatial data. Recently, there has been significant interest in developing and using standard data models for municipal infrastructure assets.

This paper discusses the issue of interoperability of municipal infrastructure asset management systems. It highlights the importance of data standards and the data requirements for interoperability, and describes some of the challenges of maintaining municipal asset data in a form that enables efficient access, retrieval, and sharing. It then provides a broad survey of several efforts to develop and implement these standards. Finally, the paper discusses an ongoing effort to develop a standard data model for wastewater systems. A prototype implementation that demonstrates the interoperability of geographic information systems (GIS) and hydraulic modeling software for a typical sewer network is also presented.

2. Importance of data modeling standards for infrastructure asset management applications

Exchanging data between different asset management systems is usually achieved in two ways:

- Special-purpose tools translate the data between different vendor-specific data models and formats. The translation typically involves a significant amount of data retrieval, reinterpretation, and reentry by end users. The “application-specific" nature of the translators also imposes unnecessary constraints by requiring the use of proprietary vendor-specific data models and software systems.

- Application programming interfaces (APIs) or software components are used for inputting or extracting data directly to and from different vendor-specific applications. Many commercial applications offer APIs so that other applications can access their internal data model. Although some of these APIs appear to be de facto industry standards, such as the component object model (COM) and open database connectivity (ODBC), the majority of the APIs are proprietary and application specific.

The inherent limitations of these two methods have created a consensus among the industry and research community that vendor-neutral standard data models and formats constitute the most viable option for software interoperability. In addition, standard data models would enhance the operational efficiency and asset management cost-effectiveness in municipalities in many ways.

First, standardizing the semantics and format of a data model can ensure the integrity and consistency of the data across various disciplines and throughout the life cycle of municipal assets. The standard data model can provide common definitions of the asset data (e.g., entities, attributes, and data quality metrics), which can promote consistency and enable the effective use and management of data. When a data model is standardized and endorsed by major bodies in the industry, software vendors make their products and solutions compatible with this standard, which further ensures data consistency between systems. Consultants and contractors can then be required to make their data compatible with the standard, making the data easily accessible and reusable for downstream maintenance and operations management.

Second, a standard data model can enable municipalities to streamline and coordinate their workflow processes by integrating data across different departments, improving data flow, and sharing data within and across their organizational boundaries (e.g., with other utility companies). Different stakeholders can access and exchange data in a common and consistent manner. Data could be reused, and the duplication and potential inconsistencies in collecting, validating, and managing data could be reduced or eliminated.

Third, a standard model can enable data sharing between municipal asset management systems and other enterprise-wide systems (e.g., enterprise resource planning systems). It can support the efficient exchange of spatial and nonspatial data between disparate and heterogeneous software systems. Municipalities can integrate various aspects of the asset data into a unified data model to enable the seamless interoperation of legacy asset management applications (Halfawy et al. 2002).

Fourth, the life cycle of municipal data is extremely long compared with that of any data model or software technology. Software systems and proprietary data formats typically go through extensive changes and upgrades in short periods of time. Special vendor-specific software may be needed to translate older data models or formats to newer ones after just a few years. An open, standard, and neutral data format is one way to ensure that the data created and modified (at high cost to the organization) can survive changes in software and data technologies. Unlike proprietary data models, which are usually owned by a few predominant vendors, open data standards are owned by and available to all parties. The data will not be "locked" by proprietary software systems, and users will have more control over their own data. The data can continue to be retrieved and reused even if the software becomes outdated.

Finally, most municipalities lack a formal and documented hand-over strategy for design, construction, and rehabilitation information at the end of projects (e.g., as-built drawings and specifications), and they lack a way for integrating this information back into their databases to reflect the as-built status of the assets. The existing paper-based data exchange methods may result in data loss and inconsistencies, as well as the inability to reuse the data later. A standard data model can help municipalities develop and implement a strategy that supports effective handover protocols and allows for the reusability of data throughout the life cycle of infrastructure assets. The model can also be used to define the content and format of the information to be handed over.

3. Spatial data modeling standards

Infrastructure assets are generally defined by their spatial and nonspatial data. Spatial data consist of the geometric and positional attributes of the assets and provide a vital support for many of the activities and decision-making processes of most municipal information systems. Nonspatial data, on the other hand, include all other aspects such as the physical, functional, and performance attributes of the infrastructure assets, as well as the operational, maintenance, rehabilitation, and cost attributes. The majority of nonspatial data can typically be linked to, or associated with, some spatial attributes. The ability to integrate spatial and nonspatial asset data would significantly enhance the efficiency of data access, management, querying, and analysis.

Recently, there has been a growing realization that spatial data models and GIS can play an important role in supporting municipal asset management (Halfawy et al. 2000; Vanier 2004). As a result, there has been a trend to augment existing asset management systems with GIS functionality. Several legacy asset management systems have already implemented GIS functionality internally or supported links with commercial GIS. A major advantage of using GIS in municipal applications is its capability to integrate data sets from several distributed and possibly heterogeneous systems and data sources. For example, a complete spatial representation, showing the land, infrastructure, and utilities information of a specific municipality, can be achieved by overlaying different layers (or themes) on a base map. These themes can exist and be maintained across several departments or organizations.

Since a substantial portion of infrastructure asset data is spatially related, the use of available spatial data standards will clearly facilitate the process of defining comprehensive and integrated data models for municipal infrastructure assets. During the past decade, there has been a surge in activities to standardize spatial data models and formats, primarily across various GIS software. The rest of this section discusses two predominant spatial data standards that can potentially support the development of integrated data models for municipal assets: the Open GIS Consortium (OGC) standards and the International Organization for Standardization, Technical Committee 211 (ISO/TC 211), standards.

3.1. Open GIS Consortium standards

The OGC standards aim "to provide a single 'universal' spatio-temporal data and process model that will cover all existing and potential spatio-temporal applications; to provide a specification for each of the major database languages to implement the OGIS [open geodata interoperability specification] data model; and to provide a specification for each of the major distributed computing environments to implement the OGIS process model" (OGC 2005). The OGC approach to interoperability involves defining standard software interfaces, rather than using a standard data model and neutral file format. These interfaces support open and standardized access to distributed spatial data and processes, enabling interoperability and integration of spatial data resources across the Internet. The OGC standards are widely accepted in the industry and are already supported by many software vendors.

The OGC has developed two main sets of standards: abstract and implementation specifications. The abstract specifications define conceptual data models and interfaces that are used to develop the more detailed implementation specifications. Implementation specifications, on the other hand, are detailed descriptions of APIs that are based on the abstract standards. The OGC standards also define a set of services and metadata standards that enable the discovery and access of spatial data and services across the Internet. The OGC reference architecture has adopted a Web services model, based on the extensible markup language (XML), for integration and interoperability of different spatial data resources and data processing systems.

The OGC abstract specifications are organized into topics and can generally be categorized into two main groups. The first group focuses on standardizing the data models, while the second group focuses on specifying APIs of Web services to enable interoperability of distributed spatial systems in a plug-and-play fashion. The OGC implementation specifications describe the schemas for Web services interfaces and the geographic markup language (GML). GML provides a set of XML-based schemas and a generic framework for defining domain-specific applications schemas. GML also provides the mechanisms and conventions for modeling, encoding, and exchanging the spatial and nonspatial attributes of objects. Examples include geometry for spatial features, coordinate reference systems, topology, and coverage schemas. GML is becoming a dominant standard for spatial data modeling and exchange.

3.2. ISO/TC 211 Geographic information/Geomatics standards

The International Organization for Standardization (ISO) Technical Committee (TC) 211 is a fairly recent initiative that includes representatives from 33 participating countries, 17 observing countries, and many external observer organizations. The standards define spatial data models and processes for managing, acquiring, processing, exchanging, analyzing, and presenting spatial data in a consistent and standardized form (ISO/TC 211 2005). ISO/TC 211 currently has approximately 40 standards in development. Nine international standards and 10 draft international standards have been released, and the remaining ones are in preparation. A complete list of these standards can be accessed on the committee's Web site (ISO/TC 211 2005).

The most important ISO/TC 211 standards for infrastructure management applications are ISO 19107 (spatial schema), ISO 19115 (metadata), ISO 19111 (spatial referencing by coordinates), and ISO 19112 (spatial referencing by geographic identifiers). ISO/TC 211 data models are defined using the unified modeling language (UML).

ISO/TC 211 has adopted a number of existing OGC specifications. Most notable is the OGC's spatial schema, *Abstract Specification Topic 1: Feature Geometry*, which was adopted as ISO 19107. The schema describes the spatial attributes of geographic features and a set of operations for data access, querying, exchange, and management. The OGC service architecture was adopted as ISO 19119. GML, currently being developed jointly by OGC and ISO/TC 211, is to be released as ISO 19136. A number of other OGC specifications, such as the Web mapping services and the Web feature services, have also been submitted to ISO/TC 211 for consideration.

4. Data modeling standards for municipal infrastructure systems

During the past decade, and parallel to the efforts to standardize GIS spatial data models, several researchers and software vendors have been developing data models for municipal asset management applications in various domains. Many of these models have now reached a level of maturity that would enable them to be widely used and adopted as standard data models in the industry. This section provides an overview of the most important data standards related to municipal infrastructure assets.

4.1. Federal Geographic Data Committee data standards

The Federal Geographic Data Committee (FGDC) has developed spatial data standards to enable interoperability between different municipal GIS software (FGDC 2005*a*). Nineteen of these standards have been adopted by almost all organizations within the US Department of Defense. Many other federal and local government organizations, municipalities, and utility companies have adopted the FGDC data standards as well. The FGDC defines three main sets of data standards: data content standards, data transfer standards, and metadata standards.

4.1.1 Data content standards

The FGDC standard data models were primarily defined to support areas of facilities, municipal, and civil infrastructure management. The data content standards describe the semantics of a set of domain objects, along with their attributes and interrelationships. A data content standard in a specific domain describes the structure and content of a feature attribute table (FAT). A typical FAT includes a comprehensive listing of objects (or features) and their associated spatial and nonspatial attributes within that domain. The content standards also define a number of constraints to ensure data integrity and consistency. For municipal infrastructure applications, the most important data content standards are the cadastral content standard (FGDC-STD-003) and the utilities content standard (FGDC-STD-010).

The utilities content standard aims at standardizing the spatial information of utilities systems for engineering and maintenance management applications (FGDC 2005*b*). The standard defines utility system components by specifying the names and description of feature types and their spatial and nonspatial attributes and the domain (i.e., values range or list) of various attributes. The standard describes 11 feature classes: water distribution, wastewater collection, storm drainage collection, saltwater, natural gas distribution, compressed air, electrical distribution, electrical monitoring and control, fuel distribution, heating–cooling systems, and industrial waste. The standard also incorporates several modeling concepts, such as grouping utility system components into feature classes.

The data model defines attributes to describe the structural condition of the asset. The allowable values are: condemned, cracked, damaged, heavy damage (useable), light damage (useable), moderate damage (useable), heavy damage (unusable), light damage (unusable), moderate damage (unusable), dangerous to use, good (not new), newly built, newly built (not finished), poor, to be determined, unknown, useable, and not useable.

The FGDC data models are defined using UML and are intended to be implementation neutral. Software implementers can customize or modify these models and use any software technology or platform to develop their systems in ways that best suit their particular purpose and requirements. Although UML serves as a robust and rich data modeling method, UML models are not used for encoding and exchanging data. Therefore, a data encoding and exchange format, known as the spatial data transfer standard (SDTS) and described in the following section, is specified, along with the data content standard.

4.1.2. Spatial data transfer standard

The SDTS defines methods and a data format for representing and exchanging spatial data in a vendor-neutral manner (USGS 2005). The SDTS specifications are organized into two sets: the base specifications (SDTS parts 1–3) and the profile specifications (SDTS parts 4–7 and potentially others). The base specifications describe the spatial data model structure and content and a format for exchanging spatial data, whereas the profile specifications define the rules and formats for applying SDTS (i.e., the base specification) to the exchange of particular types of spatial data (vector, raster, etc.). Many commercial GIS applications have already implemented support for SDTS. In general, a software system supports SDTS by providing users with special commands for importing and exporting data in the SDTS format.

Part 1 of the SDTS base specifications defines a conceptual model for spatial objects (or features) that includes their geometric and topological attributes and interrelationships. It organizes the spatial data into a set of modules, and each module has a number of fields that define a specific type of spatial information. Overall, the standard defines the structure and content of 34 modules grouped into five areas: global, spatial objects, attributes, data quality, and graphic representation. A number of options related to projections, coordinate systems, and data quality are also defined. The data model defines 32 spatial object types that include point features (0D), line features (1D), and area features (2D) as vector and raster objects. It also defines components of a data quality report and the layout of SDTS modules that contain the information needed for a spatial data transfer.

Part 2 defines a standardized list of spatial features (or entity types) and their associated attributes. The features are defined using the 32 spatial object types described in part 1.

Part 3 defines the rules and formats for encoding the data model, using the ISO 8211 standard, into a neutral and structured file format that can be exchanged in text or binary format. It also specifies the method for mapping logical constructs (e.g., spatial features, modules, and fields) into the ISO 8211 data descriptive file.

The profile specifications (parts 4 and up) address the use of the SDTS standard for exchanging different types of spatial data or profiles. A profile defines rules to "restrict" the mapping of SDTS spatial data described in the base specification to a specific data type. Existing profiles include topological vector, raster, computer-aided drafting and design (CADD), nontopological vector, object-based, cadastral, graphic, geodetic point, and digital geographic exchange standard – vector product format. A profile specifies (1) the restrictions on the use of specific spatial object types; (2) the conventions for naming modules and files; (3) the options for mapping objects, such as projections and coordinate systems; and (4) the use of the ISO 8211 standard for textual or binary encoding. Software implementing support for the SDTS standard typically supports one or more of the defined profiles.

The most commonly used profiles for implementing the SDTS are parts 4 and 5. Part 4 is the topological vector profile (TVP) for exchanging vector-based spatial data that can be described as a planar graph. Part 5 is the raster profile for exchanging raster-based data that can be represented using the ISO basic image interchange format (BIIF) or the georeferenced tagged information file format (GeoTIFF). An SDTS profile can also be used to support several data formats and software. For example, the TVP can handle many data formats, including the digital line graph (DLG) format, developed by the United States Geological Survey; the topologically integrated geographic encoding and referencing format (TIGER), developed by the US Census Bureau; and the coverage format, developed by the Environmental Systems Research Institute (ESRI).

4.2. Spatial data standards for facilities, infrastructure, and environment

Developed by the CADD/GIS Technology Center, the tri-service spatial data standards and tri-service facility management standards define standard data models for infrastructure and facilities engineering and management applications (CADD/GIS 2005). These standards have been under development since 1992, and more recently, as a result of the merger of the FGDC Facilities Working Group and the CADD/GIS Technology Center, many of the FGDC data content model standards were harmonized with the spatial data standards for facilities, infrastructure, and environment

(SDSFIE). This integration of standards will eventually advance their adoption and implementation.

The SDSFIE was recently approved as a national standard by the US National Committee for Information Technology Standards (NCITS) and is now known as NCITS 353. SDSFIE has already been widely adopted by several US federal, state, and local government organizations, as well as municipalities and utility companies. Many software systems (e.g., ESRI's ArcGIS®, Intergraph's GeoMedia®, Autodesk's AutoCAD Map®, and Bentley's GeoGraphics®) have implemented the standard. The latest version of SDSFIE has added support to several new application areas, such as operational, environmental, organizational, and financial management.

The SDSFIE provides a data model for defining spatial features, such as buildings and utilities, and for describing a relational database schema for storing the attribute data associated with these features. Attribute data for a particular feature are defined in an "attribute table" attached to that feature. However, the standards do not provide implementation-level specifications or define a neutral data format for exchanging data and enabling interoperability.

The SDSFIE data model is organized into a hierarchy of entity sets, classes, and types. An entity set represents a project-level grouping of entity classes that, in turn, contain groups of entity types (or features). Each entity type is represented graphically on a map and is linked to an attribute table that contains specific information about the entity. Release 2.1 of SDSFIE defines 26 entity sets. The "utilities" entity set, for example, contains 14 entity classes, 2 of which are water and wastewater systems. The "water" entity class defines 27 entity types including the "water_line," which defines 56 attributes describing the inventory and performance characteristics of a typical water pipeline. At the implementation level in a typical GIS, an entity class would correspond to a feature data set (as in ArcGIS) or to a drawing file (as in AutoCAD). An entity type, on the other hand, would correspond to a CAD layer or a GIS theme.

4.3. Environmental Systems Research Institute data models

The ESRI has defined application-specific data models in 24 domains, including water and energy utilities, land parcels, and transportation networks (ESRI 2005). These models have been developed over many years in collaboration with users in different disciplines. As is typical for vendor-specific data models, ESRI models are primarily intended to facilitate and expedite the implementation of GIS software using the ESRI toolset (ArcGIS®). However, as ESRI has the biggest market share of GIS installations in the world (GITA 2003), the company can have a positive influence on the standardization of data models within individual disciplines.

The water utilities data model is one of the most mature models defined by ESRI. It can be customized and configured for different water utilities including potable water, wastewater, and stormwater utilities. The data model provides a comprehensive object-oriented network-wide representation of water and wastewater infrastructure systems. It defines detailed physical parameters of pipelines, including pipe location, material, diameter and depth, exterior coating, joint types, lining type, roughness, date of installation, type of pressurization, type of valve, work order administration, and general information related to the pipe network. The model also defines a number of performance parameters, such as flow measurements and pressure rating. The data model allows the building of network topologies that explicitly define and maintain connectivity and spatial relationships between the various network components. Although the ESRI data models are popular with practitioners, they generally suffer from the following limitations: (1) inspection and condition assessment data are not represented and (2) the number of attribute data defined in the model is relatively limited compared with that defined by SDSFIE.

4.4. LandXML

The LandXML initiative (LandXML 2005) is one of the earliest efforts in the industry to advocate and develop XML-based data models for infrastructure management applications. Many data modeling efforts are starting to use the World Wide Web Consortium XML schema (W3C 2005) for data modeling, encoding, and exchange. The new XML schema standards have proved to be particularly useful in supporting the development of semantic-rich object-oriented data models.

The LandXML schema defines entities for land, roads, wastewater systems, bridges, utilities, and cadastral data management applications. The current schema (version 1.1) includes a design data model and a surveying data model. The design data model covers entities such as roadways, alignments (e.g., road centerline, horizontal and vertical alignment curves, and cross sections), a 3-D terrain model, and pipelines networks. The survey data model includes entities such as coordinate geometry point elements, coordinate systems, land parcels, and raw data collection parameters and measurements from surveys.

The current LandXML schema seems to focus primarily on defining land and road classes. Although the schema defines a limited number of classes for pipe networks, the schema has been mostly dedicated to describing hydraulic parameters, such as the pipe flow and hydraulic grade attributes. The number of attributes in the schema is extremely limited, so the model is not practical for water and wastewater applications. Significant work is still needed to extend the model into these areas. Moreover, the schema does not define attributes related to condition or maintenance data.

4.5. Municipal infrastructure data standard

The municipal infrastructure data standard (MIDS) has been developed by Ontario's Tri-Committee for the Utilization of Information Technology in Public Works (Tri-Com 2005). MIDS is a joint activity of the Ontario Good Roads Association, the Municipal Engineers Association, and the Ontario Public Works Association (a chapter of the American Public Works Association. The standard aims to integrate and extend a number of previously defined data models for water, wastewater, roads, and bridge infrastructure systems for the Province of Ontario. Unlike other efforts, which are primarily supported by research-oriented organizations or technology providers, MIDS was initiated by practitioners to support interoperability through standardized data models and formats.

The MIDS specifications are primarily intended for use by infrastructure owners to support the efficient collection and management of infrastructure life-cycle data. A number of standard data models have been defined for roads, water distribution networks, wastewater and storm collection networks, and structures, among others. In addition to defining data content standards, MIDS defines "business rules" standards to specify the workflow processes required to maintain the integrity and validity of the data.

One important characteristic of MIDS is its focus on defining asset inspection and condition data. However, MIDS is still under development, and the data models are not available in the public domain.

4.6. Pipeline open data standard and ISO 15926-2

Data models developed in the oil and gas industry can provide insight to model developers in the water and wastewater industry. As a topological network of pipelines, nodes, and structures, oil and gas pipeline data models and water and wastewater data models share many characteristics. Standard data models for oil and gas pipelines have been widely accepted and adopted for use in the industry. These models can provide useful guidelines on modeling requirements and schemes for representing nonspatial attributes and integrating these attributes into spatial data models to obtain more comprehensive and integrated schemas. Two such models deserve special attention.

The pipeline open data standard (PODS) defines a relational data model to describe networks of oil and gas pipelines (PODS Association, Inc. 2005). The PODS data model is distinguished by a large number of attributes related to inspection, condition, risk assessment, maintenance, and work management. No other data model addresses these aspects to the same level of detail. Many of these attributes could be easily adapted for use by the water and wastewater data models.

ISO 15926-2 is another related data standard in the oil and gas industry; it integrates the life-cycle data for production facilities throughout various phases of engineering, construction, operation, and maintenance (ISO 2005). This standard comprises two main components: the European Process Industries STEP [standard for the exchange of product model data] Technical Liaison Executive (EPISTLE) core data model and the EPISTLE reference data library (RDL). The EPISTLE core data model represents the life-cycle information of the process plant, whereas the RDL represents a collection of information about classes common to a number of users, process plants, or both. Supporting a typical life-cycle activity would involve the use of the reference data in conjunction with the data model. More specifically, this standard deals with many types of process plants (oil, gas, electrical, steam, and others); associated structures; and the installation, commissioning, maintenance, and replacement of equipment.

4.7. Industry foundation classes standard

The majority of existing infrastructure data standards focus on modeling "linear" municipal assets and, therefore, do not adequately address the modeling of nonlinear assets, such as facilities. During the past decade, several data standards have been developed to support facilities design, construction, and management. Most notable of these is the International Alliance for Interoperability (IAI) industry foundation classes (IFC) standard (IAI 2005*a*), recently endorsed by the ISO as ISO/PAS 16739 (publicly available specification). The IFC standard represents one of the largest scale and most mature efforts to standardize facilities design and construction data. Based on ISO 10303 STEP, the IFC standard was initiated in 1994 and has since undergone four major releases, the latest of which is release 2, edition 2 (known as IFC 2x2). Many software applications have been developed to implement the IFC schema to support facilities design and construction projects (Halfawy and Froese 2005).

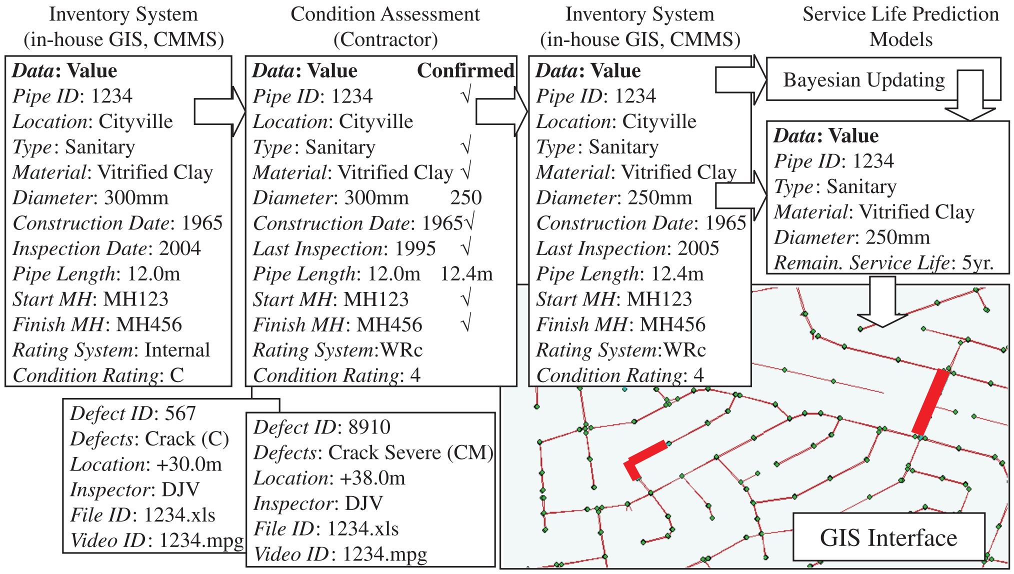
The IFC model defines a multilayer, integrated schema that represents the structure and organization of data in the form of a class hierarchy. The hierarchy covers the core project information such as building elements, the geometry and material properties of building products, project costs, schedules, and organizations. The latest IFC release has significantly extended the core model to support aspect models, such as structural steel, reinforced concrete, precast concrete, and structural analysis. A number of efforts to extend IFC to support other infrastructure projects, such as bridges, are underway.

Modeling buildings in their geographic context with respect to other municipal infrastructure assets (e.g., water, sewer, and road networks) would require the integration of GIS data into buildings data models, as well as the representation of buildings in GIS maps. Relevant GIS data may include site topology and terrain model, parcel and zoning information, water and sewer distribution systems and service connections, and access roads. Full information on buildings (not just their footprint) can be integrated into and accessed from GIS maps as geographic features. An ongoing IAI project is pursuing this objective by extending the IFC data model to integrate building information with relevant geographic information. The project, called IFC for GIS (IAI 2005*b*), defines methods for representing GIS data in IFC. This integration is required to support applications such as surveying, planning, regulations evaluation, and linking of building objects and related geographic features. The new integrated schema will be part of the next IFC release.

5. Developing integrated data models for municipal infrastructure systems

From the previous discussion, it is evident that there have been significant efforts to develop standard data models for municipal infrastructure assets. Many data models have been proposed; some have been in use for a number of years and thus have reached a level of maturity that makes it practical to adopt and deploy them. However, most of these models primarily focus on the spatial and physical aspects of infrastructure assets, and little attention has been given to modeling nonspatial life-cycle aspects, such as condition, performance, maintenance, rehabilitation, and cost data. Many of these models, although rich in detail, have not defined implementation or data format standards, and it is left up to software vendors that support the standards to define data exchange formats. However, recent advances in data modeling, especially the development of XML standards and technologies, provide new opportunities for developing more

Fig. 2. A subset and use case of the proposed data model for wastewater systems.



comprehensive and integrated data models, as well as more efficient data encoding and exchange mechanisms.

Developing a standard data model is a formidable and time-consuming task requiring the collaboration of many stakeholders. The data model must efficiently capture various aspects of each asset, including physical, functional, and performance data, as well as operational, maintenance, and cost data; and it needs to be endorsed, adopted, and supported by both practitioners and software vendors.

As part of an ongoing effort to standardize data models for municipal infrastructure systems, the Municipal Infrastructure Investment Planning (MIIP) project team, led by the Institute for Research in Construction (part of the National Research Council of Canada (NRC)), in collaboration with industrial partners, is working to harmonize and capitalize on the data models discussed above and to extend these models to incorporate various life-cycle aspects of municipal infrastructure assets (NRC 2005). The MIIP data models aim to help municipalities realize the full potential of geospatial and data management technologies to support the sustainable management of their infrastructure assets.

5.1. Wastewater collection system data model

The first phase of the data model is limited to wastewater collection systems. A sample subset of the model structure and an example use case are depicted in Fig. 2. The figure shows the main data fields needed to characterize an inventory of the system and its inspection and condition attributes. The inspection and condition attributes are based primarily on the Water Research Centre standard (WRc 2001). The inspection data include the identification and scoring of pipe defects. The spatial and inventory attributes are defined as a subset of the attributes in the SDSFIE and ESRI data models as well as in the database schemas currently used by several Canadian municipalities. Financial and administrative data also need to be integrated, but they are outside the scope of the project at this time.

In a typical use case of the data model shown in Fig. 2, the inventory and inspection data can be readily transferred to a condition assessment application (step 1), at which point they are augmented by the condition rating data. Some inventory data can also be validated, updated, and returned to the inventory system (step 2). The service life prediction model can then be used to calculate the remaining life of individual sewer segments on the basis of deterioration curves developed for pipes of similar type, material, diameter, and location (step 3). Bayesian updating can be used to update these deterioration curves when more recent condition assessment data become available (step 4).

5.2. GML-based schema encoding

The wastewater integrated data model has been encoded to be compliant with the new OGC GML 3.1 schema, which will be released as ISO 19136 (see section 3.2). The proposed data model uses the GML schema elements and data types to encode the spatial and nonspatial data for various components of the wastewater system. New domain-specific features are defined as subtypes of *gml:AbstractFeatureType*, which defines the basic model for geographic features. GML requires that all application-specific feature types be derived from *gml:AbstractFeatureType*, and be defined as immediate child elements of the global <*schema*> element.

The following excerpt of the XML schema shows an enumeration of sewer material types and an XML *complexType* of the sewer pipe feature.

```
<xsd:simpleType name="SewerMaterialEnum">
  <xsd:restriction base="xsd:string">
    <xsd:enumeration value="Brick"/>
    <xsd:enumeration value="Concrete"/>
    <xsd:enumeration value="VitrifiedClay"/>
    <xsd:enumeration value="PVC"/>
    <xsd:enumeration value="Others"/>
    <xsd:enumeration value="UserDefined"/>
    <xsd:enumeration value="NotDefined"/>
  </xsd:restriction>
</xsd:simpleType>

<xsd:simpleType name="SewerTypeEnum">
  <xsd:restriction base="xsd:string">
```

```
    <xsd:enumeration value="GravityMain"/>
    <xsd:enumeration value="ForceMain"/>
  </xsd:restriction>
</xsd:simpleType>
<xsd:complexType name="SewerPipe">
<xsd:complexContent>
  <xsd:extension base="gml:AbstractFeatureType">
  <xsd:sequence>
  <!-- child elements of the sewer pipeline type -->
    <xsd:element name="PipeID" type="xsd:ID"/>
    <xsd:element name="Material" type="SewerMaterialEnum"/>
    <xsd:element name="Type" type="SewerTypeEnum"/>
    <xsd:element name="Diameter" type="xsd:double"/>
    <xsd:element name="Thickness" type="xsd:double"/>
    <xsd:element name="Length" type="xsd:double"/>
    < xsd:element name="BuriedDepth" type="xsd:double"/ >
    <xsd:element name="UpStreamInvertLevel" type=
"xsd:double"/>
    <xsd:element name="DownStreamInvertLevel" type=
"xsd:double"/>
    <xsd:element name="UpStreamManholeID" type=
"xsd:IDREF"/>
    <xsd:element name="DownStreamManholeID" type=
"xsd:IDREF"/>
    <xsd:element name="InstallationDate" type="xsd:date"/>
    <xsd:element name="LastInspectionDate" type="xsd:date"/>
    <xsd:element name="ConditionRatingProtocol" type=
"ConditionRatingProtocolEnum"/>
    <xsd:element name="ConditionRating" type=
"ConditionStateEnum"/>
    ….
  ….
  </xsd:sequence>
</xsd:extension>
</xsd:complexContent>
```

The *SewerPipe* features will inherit all the *gml:AbstractFeatureType* properties (e.g., *gml:location* and *gml:boundedBy* properties), as well as the properties of its parent super types (e.g., the *gml:id*, *gml:metaDataProperty*, *gml:description*, and *gml:name* properties inherited from *gml:AbstractGMLType*).

6. Interoperability example: sharing data between GIS and hydraulic modeling systems

A prototype GIS-based sewer management system was developed to support the modeling and maintenance management of sewer networks (Halfawy et al. 2000). This software demonstrates the potential benefits of adopting a unified and standard integrated data model to enable efficient data exchange and interoperability. A more detailed description of the software can be found in Halfawy et al. (2000).

The GIS software component was developed using ESRI's ArcView® software. The component stores and manages a spatially based representation of the sewer network inventory data, which includes the structural characteristics of the sewer system (e.g., geometry, material, classification, materials, and thickness) and maintenance records. Users can use the GIS component to access and update the sewer system data and to navigate through the information by accessing cross-referenced data.

The sewer system condition data were collected in an extensive sewer system evaluation survey (SSES). The survey employed many techniques including visual inspection, closed circuit television (CCTV) inspections, smoke testing, and dyed-water testing. The SSES also identified and investigated performance problems that typically arise as a result of sewer deterioration (e.g., infiltration–inflow and structural deficiencies).

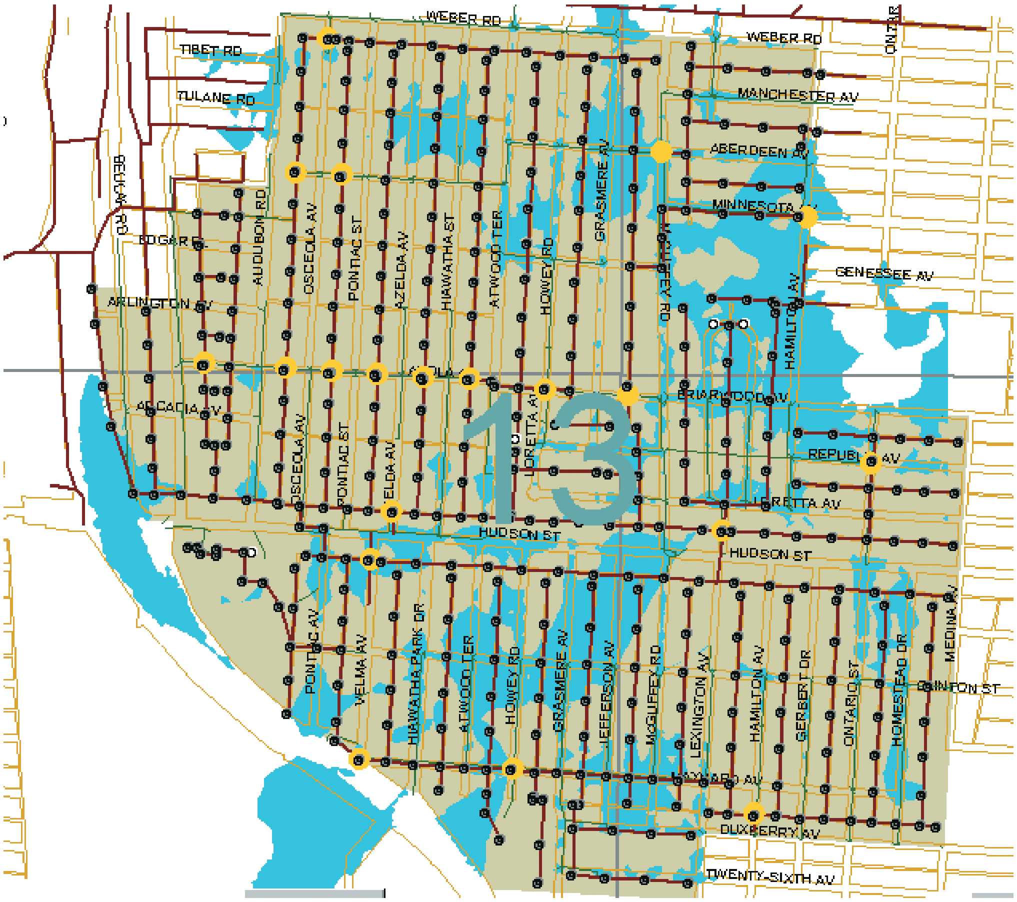
The defined data model represented a comprehensive view of the sewer network's structural, spatial, condition, and performance characteristics, with links to a wide range of "unstructured" information. Unstructured information includes inspection reports, video clips of the CCTV inspection, manhole and sewer pictures, and CAD drawings. Two relational database management systems were developed to store and manage the sewer system data. The first database stores all the inventory data, which includes all the details of the sewer system's physical characteristics. It was developed using the Hansen® Infrastructure Management System, which interfaces with an Oracle® database management system. The second database stores the sewer system condition data collected from the SSES and was developed using Microsoft® Access™.

The GIS interface enables users to access, query, and manipulate the sewer system data by simply pointing to map features, which represent the sewer system components. For example, a manhole is linked to its inspection reports, images of the manhole, CAD drawings, and results of field tests performed at the manhole or at a location close to the manhole.

The hydraulic modeling and simulation component were encapsulated with the US Environmental Protection Agency's Storm Water Management Model (SWMM) software. This component enables the user to analyze the response of the sewer system under various weather conditions, to predict and assess the consequences of various events (e.g., flooding or surcharge that could occur as a result of heavy rain), to evaluate different maintenance scenarios, and to perform "what-if" analyses. Preparing SWMM data files and interpreting the output files can be a laborious process, which limits the user's ability to conduct in-depth simulations and performance analyses. As a result, users typically create simplified models and limit their investigation to a few what-if scenarios. Although the GIS database and the hydraulic model share a considerable amount of data, these systems have traditionally been used separately. This has limited the user's ability to integrate these tools to support operational and maintenance decisions.

With the definition of an integrated data model, the construction of the hydraulic models was largely automated, as the sewer system data could be extracted directly from the GIS database. After running the simulation, the hydraulic model generates output files in a text-based form that includes a tabular description of flow within the modeled sewer network (e.g., information about maximum flow levels and pipe capacity). Interpretation of these output files, especially large ones, not only is a tedious process, but also requires extensive expertise and familiarity with the software and its data format. However, integration of the simulation

Fig. 3. Using an integrated data model to share data between GIS and hydraulic modeling software (from Halfawy et al. 2000).



results with the GIS database enables users to visualize and get insight into the enormous amount of data generated from the simulation runs. For example, if the simulation results indicated that a flooding event occurs, then a simple query can highlight the problem areas on a map, as shown in Fig. 3. The interoperability of the GIS and hydraulic modeling software were proved to significantly improve the user's ability to evaluate the sewer network performance using several what-if scenarios. Without such data sharing and interoperability, these evaluations would be prohibitively costly and time consuming.

7. Conclusion and future directions

Inefficient data exchange and the lack of interoperability between different municipal asset management systems have been major impediments in the management of asset information. Efficient data access and retrieval supports cost-effective decision making at all levels of municipal asset management. Much inefficiency has been attributed to the use of inconsistent data models across different asset management applications. Standard data models need to be developed and adopted to improve the availability and consistency of data across different software systems and platforms, to integrate data across various disciplines, and to facilitate the flow and exchange of information between various parties involved.

This paper described ongoing work to develop standard data models for municipal wastewater systems. A number of relevant data standards and requirements of standard data models were described. In addition, the paper discussed the importance of interoperability to the efficiency with which asset life-cycle data can be shared and enhanced. The value and necessity of this work were illustrated by examples of data sharing between inventory and condition assessment systems and between GIS and hydraulic modeling systems.

Future work is needed to extend the preliminary data models developed at this stage and to augment the spatial data models with specific nonspatial data elements, in particular by linking condition and performance data to spatial data. Future work should also focus on refining and harmonizing the existing data models discussed in the paper.

Although the initial cost of developing a standard data model is significant, the long-term return on this investment can produce tremendous benefits that are critical for supporting the efficient management of municipal infrastructure assets. Municipalities need to participate in the efforts to develop standard data models to ensure that their requirements are addressed. Whenever possible, municipalities should also adopt software solutions that support existing data standards and should encourage software vendors to implement these standards.

Acknowledgments

This work was initiated as a collaborative research effort between the University of British Columbia and NRC. The authors would like to thank the anonymous reviewers for their constructive criticism and valuable feedback.

References

CADD/GIS. 2005. The CADD/GIS Technology Center [online]. Available from http://tsc.wes.army.mil [cited 12 September 2005].

ESRI. 2005. Environmental Systems Research Institute water utilities ArcGIS data models [online]. Available from http://support.esri.com/index.cfm?fa=downloads.dataModels.gateway [cited 12 September 2005].

FGDC. 2005*a*. The Federal Geographic Data Committee [online]. Available from http://www.fgdc.gov [cited 12 September 2005].

FGDC. 2005*b*. The FGDC utilities geospatial data content standard [online]. Available from http://www.fgdc.gov/standards/status/sub3_1.html [cited 12 September 2005].

GITA. 2003. The geospatial technology report 2003–2004. Geospatial Information & Technology Association (GITA), Aurora, Colo.

Halfawy, M.R. 2004. The interoperability of geographic information systems for municipal asset management applications. Institute for Research in Construction, National Research Council of Canada, Ottawa, Ont. Municipal Infrastructure Investment Planning Project Report B-5123.4.

Halfawy, M.R., and Froese, T. 2005. Integration of data and processes of AEC projects using the industry foundation classes. *In* Proceedings of the CSCE 6th Construction Specialty Conference, Toronto, Ont., 2–4 June 2005. Canadian Society for Civil Engineering, Montréal, Que. [CD-ROM]. Paper no. CT-175. Also available online from http://irc.nrc-cnrc.gc.ca/fulltext/nrcc48131/ [cited 12 September 2005].

Halfawy, M.R., Pyzoha, D., Young, R., Abdel-Latif, M., Miller, R., Windham, L., and Wiegand, R. 2000. GIS-based sanitary sewer evaluation survey. *In* Proceedings of the 20th Annual ESRI International User Conference, San Diego, Calif. ESRI Press, Redlands, Calif. Also available from http://gis.esri.com/library/userconf/proc00/professional/papers/PAP158/p158.htm [cited 12 September 2005].

Halfawy, M.R., Pyzoha, D., and El-Hosseiny, T. 2002. An integrated framework for GIS-based civil infrastructure management systems. *In* Proceedings of the CSCE 2002 Conference, Montréal, Que., 5–8 June 2002. *Edited by* M.-J. Nollet, M. Trepanier, and S. Gaskin. Canadian Society for Civil Engineers, Montréal, Que. [CD-ROM]. Paper no. GE-012.

IAI. 2005*a*. Industry foundation classes—IFC 2x2 [online]. International Alliance for Interoperability, North American Chapter. Available from http://www.iai-international.org [cited 12 September 2005].

IAI. 2005*b*. Industry foundation classes for GIS (IFG) [online]. Available from http://www.iai.no/ifg/ [cited 12 September 2005].

ISO. 2005. Integration of life-cycle data for oil and gas production facilities (oil and gas). ISO 15926-2 [online]. International Organization for Standardization, Geneva, Switzerland. Available from http://www.tc184-sc4.org/SC4_Open/SC4_Work_Products_Documents/Oil_and_Gas_(15926) [cited 12 September 2005].

ISO/TC 211. 2005. ISO Geographic Information/Geomatics Technical Committee [online]. Available from http:// www.isotc211.org [cited 12 September 2005].

LandXML. 2005. The LandXML initiative [online]. Available from http://www.landxml.org [cited 12 September 2005].

NRCC. 2005. Municipal infrastructure investment planning (MIIP) [online]. Institute for Research in Construction, National Research Council Canada, Ottawa, Ont. Available from http://irc.nrc-cnrc.gc.ca/ui/bu/miip_e.html [cited 12 September 2005].

OGC, Inc. 2005. Welcome to the OGC Website [online]. Available from http://www.opengis.org [cited 12 September 2005].

PODS Association, Inc. 2005. The pipeline open data standard [online]. Available from http://pods.org [cited 12 September 2005].

Tri-Com. 2005. Municipal infrastructure data standard [online]. Tri-Committee for the Utilization of Information Technology in Public Works. Available from http://www.tricom.org/files/mids/MIDS3Overview.pdf [cited 12 September 2005].

USGS. 2005. Welcome to the spatial data transfer standard (SDTS) information site [online]. United States Geological Survey. Available from http://mcmcweb.er.usgs.gov/sdts [cited 12 September 2005].

Vanier, D.J. 2001. Why industry needs asset management tools. ASCE Journal of Computing in Civil Engineering, **15**(1): 35–43.

Vanier, D.J. 2004. Geographic information systems (GIS) as an integrated decision support tool for municipal infrastructure asset management. *In* Proceedings of the CIB 2004 Triennial Congress, Toronto, Ont., 2–9 May 2004. Institute for Research in Construction, National Research Council Canada, Ottawa, Ont. pp. 1–10. Also available from http://irc.nrc-cnrc.gc.ca/fulltext/nrcc46754 [cited 12 September 2005].

Vanier, D.J., and Rahman, S. 2004. A primer on municipal infrastructure asset management [online]. Institute for Research in Construction, National Research Council of Canada, Municipal Infrastructure Investment Planning Report. Available from http://irc.nrc-cnrc.gc.ca/pubs/fulltext/b5123.3/b5123.3.pdf [cited 12 September 2005].

W3C. 2005. XML schema [online]. Available from http://www.w3.org/XML/Schema [cited 12 September 2005].

WRc. 2001. WRc sewerage rehabilitation manual. 4th ed. Water Research Centre (Great Britain), WRc Publications, Wiltshire, UK.