INTERNATIONAL STANDARD

ISO 37166

First edition
2022-02

Smart community infrastructures — Urban data integration framework for smart city planning (SCP)

Infrastructures urbaines intelligentes — Cadre d'intégration des données urbaines pour la planification des villes intelligentes

Reference number
ISO 37166:2022(E)

Contents

Page

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 268, *Sustainable cities and communities*, Subcommittee SC 1, *Smart community infrastructures*.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

The city is a product of social evolution, technology, economic and social civilization improvements, as well as a fundamental unit for the social and economic life of its region. However, with the influence of global urbanization, increasingly more problems have been observed, such as environmental pollution, traffic congestion, insufficient resources and urban lifeline system weakness.

Urban planning refers to the conduct of engineering construction, economy, society, environment and land use of the city and its surroundings. It involves the regional layout of industry, the regional layout of buildings, the setting of transportation infrastructure and the planning of urban engineering. It is related to urban development and city infrastructure construction.

The planning, construction, operation, management and evaluation of community infrastructure is the process of natural environment transformation. This process involves multiple city managers and various data. Therefore, the integration of heterogeneous data for smart community infrastructure planning is particularly important. Based on ecological and spatial information, the smart city planning (SCP) data and infrastructure data that need to be integrated should be analysed. The establishment of a data integration framework and further realization of heterogeneous data integration is intended to support the operation of community infrastructure construction projects throughout their life cycles and ultimately achieve inclusive, sustainable and high-quality development of the city.

In terms of smart community infrastructure, ISO/TS 37151 describes the principles and requirements of performance metrics. ISO/TR 37152 gives possible issues and solutions in developing and operating smart community infrastructure, outline and benefits of a common framework for development and operation. In addition, BS/PAS 183 provides data interoperability, types of data, data protection reform, data value chain, purposes for data use, assessing data states, access rights for data and data structure.

ISO/TS 37151, ISO/TR 37152 and BS/PAS 183 provide the basis and guidance for ISO 37156, which describes the types and model, opportunities, privacy and security of data exchange and sharing, and provides guidance for data exchange and sharing of smart community infrastructure. ISO 37156 provides guidance for the integration of infrastructure data in this document, and this document is considered to be an application scenery of ISO 37156 in data integration.

Smart community infrastructures — Urban data integration framework for smart city planning (SCP)

1 Scope

This document establishes a data framework that involves possible multi-source common data through standardized data integration and sharing mechanisms. It includes recommendations for:

— precision, dimensions of the data, data collection, updates and storing mechanisms;

— a data model for data integration, data standardization and data fusion approaches for heterogeneous smart city infrastructure data;

— a data security level and sharable attributes for all involved data, principles on data sharing or exchange.

This document focuses on the integration and application of heterogeneous data from urban infrastructure systems, such as water, transport, energy, drainage and waste, so as to support smart city planning (SCP). It contains case studies, in Annex A, of various smart city projects.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at https://www.electropedia.org/

3.1
data
reinterpretable representation of information in a formalized manner suitable for communication, interpretation or processing

Note 1 to entry: Data can be processed by humans or by automatic means.

[SOURCE: ISO/IEC 2382:2015, 2121272]

3.2
data availability
property of being accessible and usable upon demand by an authorized entity

[SOURCE: ISO/IEC 27000:2018, 3.7, modified — term revised.]

3.3
data exchange
accessing, transferring, and archiving of *data* (3.1)

[SOURCE: ISO/TS 13399-5:2014, 3.7, modified — definition revised.]

3.4
data sharing
reference for providing shared, exchangeable and extensible *data* (3.1) to enable urban infrastructure service

[SOURCE: ISO 37156:2020, 3.3.6, modified — definition revised.]

3.5
data type
defined set of *data* (3.1) objects of a specified data structure and a set of permissible operations, such that these *data* (3.1) objects act as operands in the execution of any one of these operations

[SOURCE: ISO/IEC 20546:2019, 3.1.12, modified — Notes to entry removed.]

3.6
heterogeneous data integration
optimization method to enable effective and transformative use of data and technology from multi-source to support sustainable development of cities and to improve the management and control of space and resources

3.7
information
data (3.1) in context with a particular meaning

[SOURCE: ISO5127:2017, 3.1.1.16, modified — definition revised.]

3.8
information resource
asset
record
<set of data> document or item in physical or digital form that contributes to human knowledge

EXAMPLE Abstracting and indexing database.

Note 1 to entry: Information resource sometimes has a broader meaning, including information content, and also technology resources, human resources and financial resources that enable information content management.

[SOURCE: ISO 5127:2017, 3.1.1.44; modified — definition revised and Note 2 to entry added.]

3.9
life cycle
<of a system, product, service, project or other human-made entity> evolution from conception through to destruction or recycling

[SOURCE: ISO/IEC/IEEE 15288:2015, 4.1.23, modified — definition revised.]

3.10
security
condition that results from the establishment and maintenance of protective measures that ensure a state of inviolability from hostile acts or influences

[SOURCE: IEC Guide 120:2018, 3.13]

3.11
smart community infrastructure
community infrastructure with enhanced technological performance that is designed, operated and maintained to contributed to sustainable development and resilience of the community

[SOURCE: ISO 37156:2020, 3.1.4]

3.12
smart city planning
SCP
technical and political process concerned with the development and design of land use and the built environment, which are enhanced by the effective and sustainable integration of informational, physical and social systems and the transformative use of heterogeneous data and technology

3.13
system
set of interrelated or interacting functions constituted to achieve a specified objective

4 Principles

4.1 General

4.1.1 General

The data gathered and to be integrated for SCP should meet some general principles so as to ensure the validity of the following data integration process.

4.1.2 Data availability

The data to be integrated for SCP comes from various sources, some of which are private and classified prior to any data-sharing agreements. Thus, the description of shared data in terms of, for example, attributes, dimensions and volumes should be available to associated integration subjects (described in 6.2), so as to determine whether the shared data are truly available for the intended data integration purpose.

4.1.3 Sovereignty over the data

The ownership of the source data needs to be respected during the whole data integration process among associated stakeholders.

4.1.4 Data security

The data to be integrated for SCP should be secured during the data integration process, from data retrieval, data clean, data storage and data output.

Regional and national security requirements such as the EU General Data Protection Regulation shall be considered. Based on ISO/IEC 27000, considering domestic regulations and technological conditions, an information security management systems (ISMS) or alternative necessary data security procedure and tools should be introduced to deter possible hacker attacks and other misapplication.

The data exchange should therefore be kept to a minimum and a low level of detail. Security-relevant data for planning, constructing, operating and managing of infrastructures should basically remain with the data-collecting organization, ordinarily the utilities.

4.1.5 Data privacy

The data from community infrastructure to be integrated for SCP usually contains private information, from individual socio-economic characteristics to spatial-temporal behavioural data. Integrating and further analysing these individual-based data help to evaluate and optimize urban system performances. But individual privacy should be respected.

The use of source data during the whole integration process should be kept on an anonymous basis. The integration of individual data, for example consumer consumption data, should be pre-desensitized, without personal information being exposed to either data integration engineers or terminal users.

4.1.6 Co-construction and sharing

It is possible that an intended data integration requires data from different agencies or stakeholders. It is recommended that the necessary data are co-constructed and shared among them on a voluntary basis. In addition, the integration results should be shared with contributors.

4.2 Principles of heterogeneous data integration

4.2.1 General

In practice, the multi-source data used for data integration varies in format, dimensions, accuracy and durability. Although data integration approaches are evolving with more recognition or explanation power, it is not usually encouraged to spend huge amounts on retrieving missing data. Depending on the availability and intended use of data, it is recommended that the data integration framework applies the following principles of heterogeneous data integration.

4.2.2 Unambiguity

The definitions and categorizations of entities should be clear where reasonably possible and available. Categorization should be representative and mutually exclusive.

4.2.3 Scalability

Urban data integration requirements and process are continuously updating. The integrated data need not be thoroughly completed and comprehensive in the beginning, but it is recommended that the integration results are flexible and scalable.

4.2.4 Compatibility

Data standards applied in the data integration framework shall be compatible with existing major urban data standards.

4.2.5 Modularity

It is recommended that data integration input, output and approaches or algorithms are defined as modular components, so as to be used individually or in combination for different integration purposes.

In addition to unambiguity, scalability, compatibility and modularity, the urban data integration framework should also have extensibility and interpretability. Maintaining a high level of interpretability is vital during the integration process as the goal is to support the urban design and operational decisions by municipal officials, policymakers and engineer technical staff. A useful urban data integration framework should be capable of integrating heterogeneous data in an extensible (to multiple urban systems), scalable (to the growing amounts of quickly changing urban data streams) and interpretable manner (such that it can inform decision-making).

4.3 Data quality recommendations

On the basis of ISO 19157, the following recommendations are dependent on the intended use of data. For example, zonal plans can be acceptable with a tolerance of several metres but plans for individual buildings might require an absolute positional accuracy of a few centimetres. The quality recommendation of a variety of planning data database results should ensure locational accuracy, attribute accuracy, completeness, logical consistency, geographic quality and data relationality as far as the data are available and it is even possible to determine such data.

a) Locational accuracy. It is recommended that the location given in the data satisfies data requirement in terms of geographic accuracy.

b) Attribute accuracy. It is recommended that attribute values given in the data match the actual values in the real world to the extent required by the expected use. For example, no building should be classified as a pavement, but it can sometimes matter less if a building is classified as residential when it is in fact partly commercial.

c) Completeness. It is recommended that the number of missing or excess data items in the data set, in comparison to the real world, are suitable for the intended use. For example, all planning data should contain exactly the same number of plans as actually exist, but it might not be too important if a zonal plan contains slightly more or fewer building features.

d) Logical consistency. It is recommended that the data are logically consistent with the requirements, in terms of concept consistency, value domain consistency and topology consistency. For example, no plan should be described as agreed if it is only proposed; it is unlikely that a road would not have a postal address.

e) Geographic quality. It is recommended that the geometry of the data conveys the actual data correctly. For example, buildings are defined as polygons instead of a point when the intended use is to correlate retail shops with a metro station spatially; the boundaries of adjacent polygon entities coincide.

f) Data relationality. It is recommended that entities being referenced by different data sources are mapped and related to other entities through the data model defined in the data integration framework.

It is also noteworthy that low-quality data can still be relevant and used to identify or understand the urban environment and phenomena that are poorly documented, but high-quality data allows for more accurate decision-making.

5 Data of SCP on community infrastructure

5.1 General

Community infrastructure is the fundamental safeguard for residents' lives and city development. It incorporates various equipment and systems which are utilized for economic and social activities. Urban infrastructure is still promoting living and economic development, but its impact on ecosystems cannot be ignored. Technology conditions and functional loads can affect the operational efficiency of the social and economic system and also the living quality of residents. Therefore, scientific and reasonable urban infrastructure planning is essential to maintaining urban ecological security and health.

The scientific infrastructure planning of smart cities requires data support, including community infrastructure data and SCP data. SCP data are numerous and diverse and need to be reasonably described and classified. At the same time, community infrastructure data are a core part of SCP data. Further describing the composition, definition and source of community infrastructure data is particularly important for the implementation of the SCP. In addition, further clarification of the usage of infrastructure data is needed, providing support for SCP.

5.2 Usage of community infrastructure data

5.2.1 Construction project life cycle

Integrating of relevant community information resources based on the SCP data, including current situation data, planning, data, administrative approval data and spatiotemporal big data, to realize the entire life cycle of engineering construction projects of data exchange and sharing. Throughout the entire life cycle of the engineering construction project, including planning, design, construction, management and operation, the science of decision-making can be improved, and management efficiency can be improved.

5.2.2 Urban simulation

By integrating SCP data, the distribution, scale, related activities, performance indicators and other attributes of community infrastructure can be described and used in urban planning. In order to understand the internal mechanism of urban system operation, city simulation can be performed after finding out the cause of urban problems and assessing system performance.

5.2.3 Smart transportation

Infrastructure data based on transportation can support the overall control of the transportation field and the entire process of transportation planning and management. As a result, transportation systems have the ability to sense, interconnect, analyse, predict and control cities. In addition, it can fully ensure traffic safety, make use of the effectiveness of transportation infrastructure, improve the efficiency and management level of the transportation system, and serve the smooth development of public transportation and sustainable economic development.

5.2.4 Smart grid

By integrating electricity and ICT data, real-time data collection, transmission, sharing and dynamic monitoring can be achieved, which can enhance the scientific city and rationality of power consumption, such as power consumption and transmission. The smart energy infrastructure has brought great benefits to urban economic development, energy production and utilization, and environmental protection.

5.2.5 Smart environmental sanitation

By integrating the distribution of waste transfer stations, the status of waste bins and ICT data, real-time monitoring of environmental health services, such as infrastructure resource management, waste collection, transportation, disposal and separation, can be achieved. This reduces the cost of sanitation operations and supervision, improving the efficiency and quality of sanitation operations, and ensures the cleanliness and orderliness of the urban appearance.

5.3 Smart city planning (SCP) data

This subclause presents and explains data that are useful in SCP practices. The data dimension can be larger than community infrastructure data. The purpose is to clarify the location and connection of community infrastructure data in SCP data.

Table 1 describes the classification and association of SCP data. As shown in Table 1, SCP data consists of current situation data, urban planning data, administrative approval data and spatiotemporal big data. Current situation data illustrates the current objective status of a city, including natural environment data and built environment data (e.g. buildings, parks, bridges). Urban planning data are thematic data on urban internal planning. Administrative approval data refers to data generated by government administrative approval service. Spatiotemporal big data, such as traffic flow and economic flow, is a new type of data, which is very useful for urban planning. This new type of data is playing an increasingly important role in the planning, management and supervision of urban planning. Current situation data are the basic data for urban planning; furthermore, urban planning data are an important basis for administrative approval. Data generated by plan management provides guidance for city construction, management and sustainable development. Current situation data, planning data and administrative approval data are related to spatiotemporal big data, but related content gives priority to the classification of spatiotemporal big data. Spatiotemporal big data are the focus of attention and represent an important development direction of smart community infrastructure planning. Because of their high frequency and large data volume, spatiotemporal big data require unique methods for data integration.

Table 1 shows more details of SCP data and give examples of what the data are. These descriptions are not exhaustive or mutually exclusive.

Table 1 — Examples of SCP data

Type of SCP data	Data set category	Example of SCP data		
		Example data sets	Source of data	Examples of data insights
Current situation data	**Infrastructure current situation data** traffic data[a] energy data ICT data water data waste data	Energy data: power plant location power supply	Smart grid	Location or power consumption
	Natural environment data natural landscape data geological data	Natural landscape: digital orthophoto map (DOM)	Mapping	Surface undulations
	environmental data climate data luminescence	Environmental data: outdoor air quality	Sensor	Air quality levels
	soil data hydrology data biological data, landslide areas	Soil data: soil pollution soil quality	Survey	Pollution levels Strength of soil functionality
	Built environment data park data housing estate data central business district school hospital	Housing estate: building dimensions data building occupancy	Survey or design	Height or use

[a] ISO 14817-1 contains information about data concepts.

[b] ISO 19152 defines a reference land administration domain model (LADM) covering basic information-related components of land administration.

Table 1 *(continued)*

Type of SCP data	Data set category	Example of SCP data		
		Example data sets	**Source of data**	**Examples of data insights**
	Socioeconomic data demographic data economic data social data	Demographic data: population quantity Economic data: consumer price index Social data: administrative boundaries	Survey or census Statistical Survey Crawl online Mapping	Increase or decrease Changes in the price level Spatial extent
Planning data	**Infrastructure planning data** traffic planning data energy planning data ICT planning data water planning data waste planning data	Water supply planning: water pipe diameter	Water supply planning	Size
	Rural planning data village layout planning strategic planning of rural revitalization	Village layout planning: village distribution	Village layout planning	Spatial location
	Land use planning data	Land use data: type of land use floor area ratio building height	Land use planning	Nature of the land (e.g. construction land/woodland/transportation land)/construction land use intensity/height limit

[a] ISO 14817-1 contains information about data concepts.

[b] ISO 19152 defines a reference land administration domain model (LADM) covering basic information-related components of land administration.

Table 1 *(continued)*

Type of SCP data	Data set category	Example of SCP data		
		Example data sets	**Source of data**	**Examples of data insights**
	Specialized planning data river planning educational layout planning sports facilities layout planning medical layout planning green space planning ecological environment planning	Educational layout planning: school size school area school location	Educational layout planning	Spatial distribution pattern
Administrative approval data	**Infrastructure management data** traffic management data energy management data ICT management data water management data waste management data	Traffic management data: traffic incident monitoring data Water management data: administrative permit of wading construction projects	Traffic monitoring and management system River and lake resource management information system	Traffic incident location Development and utilization management status
	Land management data[b] land approval data land supply data land use management data land inspection data	Land approval data: land location land price land nature	Planning permit management system	Planning and design scheme
	Construction project management scheme design data scheme approval data engineering construction data planning acceptance data	Scheme design: building scale building use function building height building layout	Approval management system of construction project	Size or residential or commercial or high or low or spatial location

a ISO 14817-1 contains information about data concepts.

b ISO 19152 defines a reference land administration domain model (LADM) covering basic information-related components of land administration.

Table 1 *(continued)*

Type of SCP data	Data set category	Example of SCP data		
		Example data sets	**Source of data**	**Examples of data insights**
	Real property management data building registration data forest registration data arable land registration data grassland registration data	Building registration data: ownership information	Real estate registration system	Homeowner
Spatiotemporal big data	**Infrastructure big data:** traffic big data energy big data ICT big data water big data waste big data	Water flow: real-time water flow speed	Water sensor	Water volume
	Natural environment big data	Natural environment big data: real-time wind speed	Sensor	Speed
	Population flow	Population flow: phone signalling data	Base station equipment	Commuting characteristics
	Economic flow	Economic flow: consumption data	Credit card	Urban vitality

[a] ISO 14817-1 contains information about data concepts.

[b] ISO 19152 defines a reference land administration domain model (LADM) covering basic information-related components of land administration.

5.4 Community infrastructure data

5.4.1 General

Smart community infrastructure includes energy, water, transportation, ICT and waste. The data addressed in this document is related to the infrastructures and built environment elements that support the infrastructure. Smart community infrastructure data is created, captured, collected or curated from various sources of smart community infrastructure. Table 2 shows how the community infrastructure data works. These descriptions are not necessarily exhaustive or mutually exclusive.

5.4.2 Data definition

Table 2 defines the data from the five community infrastructures in 5.4.1.

Table 2 — Examples of community infrastructure data

Type of the infrastructure	Data set category	Example of infrastructure		
		Example data sets	Source of data	Examples of data insights
ICT	Telecommunication connectivity Internet connectivity Broadband capacity, infrastructure (gateway, transverse, radio base station), related service and license	Telecommunication line Signalling data	System logs Mobile communications operator	Megabytes of data use Communication records (latitude, longitude, time)
Energy	Electrical energy Thermal energy Fuel gas	Gas pipeline Street lighting Energy grid	Flow sensor Smart energy meters	Leaks Energy used per hour (kwh)
Waste	Industrial waste Domestic waste Solid waste	Waste bin Transfer station	Waste bin sensor Smart city management platform	Empty or full Location
Water	Water supply Rainwater Wastewater Reclaimed water Water resource	Water pipeline network Water distribution network Pump station	Flow sensor Water supply facility Smart water meters	Leaks Water load per hour
Transportation	Road Railway Water transportation Aviation Subway Public transport routes Public parking spaces Non-motorized traffic Barrier-free facility	Public shared bicycle Parking lot Barrier-free path Traffic lights	API for network transportation data Smart transportation management platform Map API	Travel path Number of parking spaces Distribution of barrier-free facilities
NOTE See ISO/TS 37151 and ISO 37156.				

5.4.3 Source of heterogeneous planning data

This subclause lists typical data sources of community infrastructures and characterizations of raw data.

City planning data may come from geographic information systems (GIS), planning and planning approval, legal texts and planning regulations. In recent years, with the development of 3D geographic information systems, 3D model data has been added. For example:

— data from geographic information systems;

— data from master plans and detailed plans;

— data from planning approval;

— data from 3D city model;

— data from sensors.

6 SCP data integration framework

6.1 General

This clause gives a data integration framework, including data integration subjects, objects, process and results. Integration subjects refer to participants during data integration. Integration objects refer to data to be integrated, which is described in Clause 5. The integration process explains the full life cycle integration activity. The integration results refer to outputs including data, tools and rules. The integration framework can be built on the basis of a unified standard, a unified base map, unified planning and a unified platform.

6.2 Integration subjects

Possible subjects of heterogeneous data integration for urban infrastructure are as follows. This list is not intended to be exhaustive.

Subject 1: Community manager, including the mayor and government industry department. Based on heterogeneous data integration for urban infrastructure, it can change the complex and inefficient situation of data sharing between government departments, promote the openness and intelligence of community development and serve community managers for scientific decision-making.

Subject 2: Data processor, including providers of data acquisition, processing and integration. The integrated data are clearer and tidier, which is easier to manage and provides more cooperation opportunities for related companies.

Subject 3: System providers, including system developers, service publishers and system operation and maintenance management. Data integration provides more data types for the system, optimizes and improves system analysis and early warning capabilities, and also provides more cooperation opportunities for related enterprises.

Subject 4: Equipment providers, including infrastructure and hardware equipment. Data integration can make infrastructure data updates more timely, effectively improve infrastructure operation and maintenance capabilities and provide more cooperation opportunities for related enterprises.

6.3 Integration objects

Integration objects refer to the data to be integrated, which is described in Clause 5 in detail. The integration of various types of data focuses on different contents.

Current situation data integration, as the basics of smart city infrastructure planning, may clarify the data source and ensure the consistency of space and time.

Planning data integration should focus on the characteristics of planning hierarchy. It is essential to ensure the consistency between different levels of planning data, such as the consistency between master planning and zoning planning, and the consistency between the special planning and the land-use planning.

Administrative approval data integration may emphasize the classification organization of metadata. Management data comes from many sources and has complex types of data, including structured data stored in the database and unstructured data such as media, audio and text. Metadata (or data labels) can be used to correlate and integrate data of different types and formats to ensure that the data can be searched when needed. For lines with gravity flow characteristics, such as rain pipes and sewage pipes, the vertical design requires additional consideration in the data integration.

Spatiotemporal big data integration may deal with high frequency and large data volume. In the process of data integration, it is necessary to filter out the real and reliable data to reduce uncertainty and realize data processing and analysis in a short time.

6.4 Integration process

Data integration needs to meet the needs of different levels, i.e establishing a reference and mapping system through data elements, metadata, classification, coding system, geospatial location information and semantic annotation. Data integration in the whole life cycle includes data acquisition, extraction, correlation, transformation and loading. It enables effective interaction and feedback among smart city infrastructures and optimizes city operation efficiency.

Spatial data are an important component of smart city infrastructure data. They represent the location, shape and distribution characteristics of infrastructure. In order to solve the heterogeneity of spatial data, data integration also needs to figure out the problems of multi-space-time, multi-scale and multi-coordinate system.

There are numerous data integration methods that aim to tackle some of the challenges of smart city infrastructure data management, such as domain-centric, data-centric and demand-centric data integration methods.

Domain-centric integration methods aim to exploit the knowledge or structure of a specific system by integrating various data streams often using network-based methods to represent urban elements as nodes and interactions between elements as edges which have largely been constrained to single system modelling. Especially for multi-system integration, new ontology-based methods have been used to embed knowledge and available data of a specific domain in an ontology (concepts or ideas and the relations connecting them to others) and to utilize the ontology to 'link' various data streams to one another.

Demand-centric integration usually employs a one-time data integration process based on a specific task. Data visualization is one form of demand-centric data integration; however, these methods often have their limitations on flexibility and scalability.

Data-centric integration methods utilize computational methods to learn the structure, properties and representation of urban data and then integrate based on such attributes. Numerous open data standards were published to manage standardized data model and exchange format based on these standards.

6.5 Integration results

The integration results mainly include integration integrated data, integration tools and integration rules. The integrated data include data in the form of files, databases through encrypted APIs, and annexes explaining input data dimensions, data systemic relationship by means of various data visualization through reports or Web services. Integration tools, which are provided for users, are integration approaches and algorithms embedded in developed integration modules or given by commercial ETL tools. Integration tools mainly include components between components and commercial ETL tools.

Visualization is essential to integration results for better understanding and using. Visualized integration data, tools and rules make it easier to understand data systemic relationships.

7 SCP data integration

7.1 General

The community infrastructure data used for SCP normally originates from different government agencies and separate public facility data systems, but sometimes have entities that conceptually overlap. Most of the time, the data are heterogeneous in terms of structure, scale and precision in time and spatial dimension, and associated with different geographic coordinate systems. The main objective of data integration is to identify and extract relevant data elements, to define universal entities, to link interrelated entities and to ensure all the data compatibly encoded.

For the smart community infrastructure data, based on a specified data model, the integration procedure normally includes data extraction, data quality verification, data encoding or mapping specification, heterogeneous data integration and any subsequent data exchange and sharing measurements, as shown in Figure 1. The data model is to define the entities of involved community infrastructures, to specify the data description and to link entities that have conceptual and/or logical relationships.

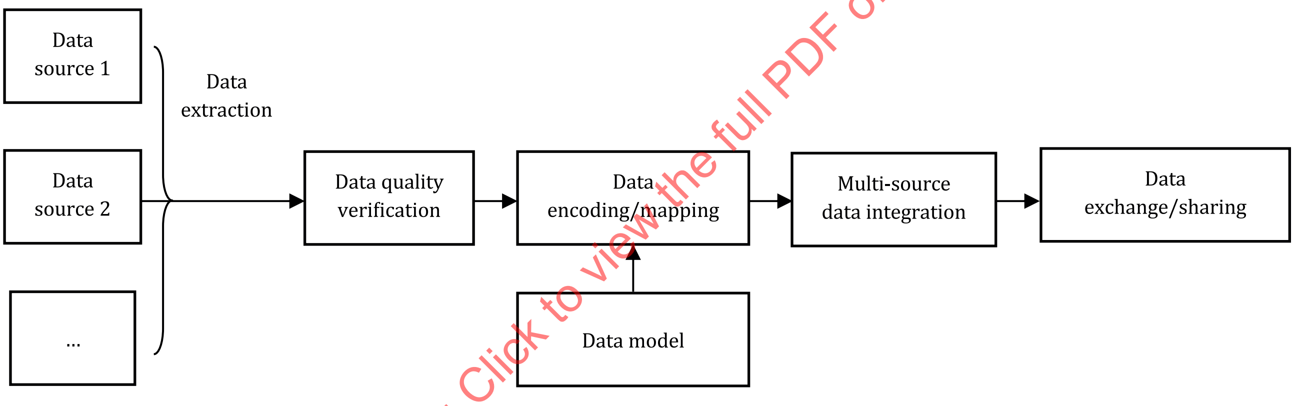

Figure 1 — Data integration procedures

7.2 Data model and description specification

The data model is built to connect among features from various data sources and give each entity a universal description specification.

The community infrastructure data used for SCP not only involve numerous correlated entities but can also be expressed as both structured and non-structured data. The widely used data models built on relational database are applicable to model most structured data. But to model numerous correlated entities and efficiently extract potential knowledge or relationship between two entities, ontology-based data models are preferable.

The description specifications of community infrastructure data entities from different data sources should be universal and characteristically described. It is recommended that the classifications of the identical entity from different data sources, such as road type, are the same or can be cross-referenced.

In addition, it is recommended that entities with geographical attributes (e.g. infrastructure boundaries, road sections, intersections) are compatibly defined and transferable with common geographic coordinate systems. ISO 19111 defines the conceptual schema for the description of spatial referencing by coordinates, and also describes the information required to change coordinates from one coordinate reference system to another.

7.3 Data extraction and system exchange

The data extracted from each data source are conferred and determined among data providers, data processors and other involved stakeholders. The data requirements can include descriptions in terms of necessary attributes, precisions, update frequency, spatial-temporal scale, data desensitization, and any particular demand specification in formatted documents. The data can be pushed or extracted through encrypted API.

The data interface between each subsystem of smart community infrastructures and the data interface with the external system uses service-oriented architecture (SOA).

7.4 Data quality verification

The data source providers should submit or offer metadata, normally along with sample data, and a data quality verification report generated by them, if any. The data integration processors should verify the extracted data quality in accordance with data requirements. A data quality verification report for each extracted piece of data is required and fed back to the data providers if necessary.

It is recommended that a suitable data quality analysis framework (DQAF) is applied for data quality verification. The dimensions and index of data quality verification should be coordinated with data integration principles.

It is also recommended that the objective and subjective data quality index and approaches are carefully chosen among data processors and possible data users.

7.5 Data encoding or mapping specification

Data encoding or mapping specification is to define the encoding or mapping criteria, such that the original data from various sources are translated and expressed in a standardised and consistent manner, before further data integration procedures.

7.6 Smart community infrastructure data entities

On the basis of ISO/TR 37150, which stipulates existing data types and relevant international and industries standards, Table 3 describes common smart community infrastructure data entities from energy, water, traffic, ICT and waste, with typical data interface, data format and data integration scenarios.

Table 3 — Smart community infrastructure data entities

Type of infrastructure	Entities (examples)	Data interface	Data format	Scenarios after data integration	Update frequency
ICT	Telecommunications line Telecoms annotation, e.g. network node label (service type, connection type, power supply type), location of the network facilities (pipeline, pole), pre-processing devices (edge or fog computing)	Data communication facilities (pipes, radio base, antenna support, power supply capacity)	Point and polyline in GIS format Text	Communication network	Monthly
Energy	Power protection line Electrical labelling Location of power station	Substation and other power supply facilities	Polyline in GIS format Text Number	Energy supply system	Monthly
Waste	Situation of the waste bin Pollutant discharge	Waste bin facilities Waste bin sensors	Point in GIS format Text Number	Waste management Pollution management	Real-time
Water	Water flow Water supply annotation	Water supply facilities Water sensor	Point and polyline in GIS format Text Number	Water mains Water supply system	Real-time
Transportation	Metro line and station Smart card swiping record	Metro facilities Smartcard corporation	Point and polyline in GIS format Text Number	Transit-oriented development	Minute
NOTE See ISO/TR 37150 and ISO 37156.					

7.7 Heterogeneous data integration

It is possible that certain smart community infrastructure data entities have multiple data sources, where each source provides partial data samples. In addition, those data samples can in some cases overlap. This generates duplicated data even after data mapping, i.e. samples from two data sources describe one object of the entity.

Therefore, it is necessary to identify possibly duplicated data, and to estimate if the data integrity satisfies data requirements; otherwise, data interpolation procedures can be incorporated.

A multi-layered data integration structure is recommended from data mapping, to entity linking, data synthesis and geospatial feature updates, by incorporating an ontology-based data model. Proper data semantic and synthesis approaches can be applied to eliminate ambiguity, to build a connection between potentially identical samples and to interpolate data.

7.8 Date management recommendations

7.8.1 General

A proper data management schema is recommended throughout the data integration process to ensure data security. It can involve the physical and/or intellectual management established over data by documenting information about their physical and logical state, their content, their provenance and their relationships with other data. The above control involves creating, capturing and maintaining information about the movement and use of documented information. The systems and processes associated with establishing control include registration, classification, indexing and tracking to ensure authenticity, reliability or non-repudiation, integrity, confidentiality, authorization and usability of data quality.

It is crucial that data access at any stage of data integration is secure and recorded from unauthorized use and illegal retrieval attempts. The EU General Data Protection Regulation (EU GDPR 2016/679) or the Data Security Management Measures of the People's Republic of China (Consultation Draft 20190528), for example, can be referred to.

7.8.2 Data exchange and sharing

Data exchange is the process of converting the structured data under the source structure into the structured data under the target mode and storing conversion details about the degree of information in terms of accuracy and completeness, so that the target data shared by involved public sectors is an accurate representation of the source data. In this context, a high level of data interoperability should be ensured.

7.8.3 Data exchange and sharing security

Security and privacy issues are not limited to residents. Service providers and governments also have their own effective reservations. It should ensure system and data security and user privacy protection. In addition, data privacy measures should consider local regulations concerning personal data processing.

8 Management of security and privacy

8.1 General

The basic suggestion of data security in cities is to guard against the corresponding security issues involved in all data source points (e.g. urban infrastructure elements, sensors) and the storage or circulation of data in urban information systems.

It is necessary to develop a security strategy which articulates the overall security policy for heterogeneous urban data integration. All agencies and organizations which are involved in providing data are responsible for maintaining safety and security.

For more details, see ISO/TR 37150 and ISO 37156.

8.2 Data security level and protection principles

In ensuring data security, data are divided into different security levels according to legal requirements, value, criticality and their sensitivity to unauthorized disclosure or modification.

The security strategy should also consider whether the security policy is justified in accordance with legislation or regulation, including voluntary participation.

Security measures should consider the sensitive data requirements (e.g. the aims and subsequent usage after the user obtains the integrated data) of personal data, intellectual property and commercially sensitive data which facilitates the provision of smart city services.

Considerations of data privacy are of equal importance to the security of data integrated. Data privacy applies to all data which is personal data or data which can be used to construct personal data about a citizen. The rights afforded to individuals under the legislation or regulation of the jurisdiction, including the right of subject access, should be considered, to ensure personal data are not transferred outside the legislative or regulated jurisdiction without adequate protection.

It is necessary that a senior management team is tasked with issuing and maintaining a privacy policy that sets a clear framework and demonstrates support for, and commitment to, the heterogeneous urban data integration. This can include managing compliance with data protection legislation and regulation, and the application of appropriate good practice policies.

8.3 Technical advice for data security

Due to the different security requirements for data at different security levels, it is recommended to focus on the technical methods for maintaining sensitive data.

The security approach should seek to preserve confidentiality, integrity and availability of data, ensuring, where possible, that data are free from danger or threat of unintended access and use. For example, damage can be caused by malware, hackers or disaffected personnel.

Any security breach of personal data should be avoided as this results in damage to citizens and the organization and potentially damages citizen trust, which can adversely impact on the whole city. It is necessary to use anonymization technology on personal data.

It is important to consider the threat from actors who seek to undermine any vulnerability in the data security measures for community infrastructure. These actors might be associated with organized crime seeking to acquire unauthorized personal or sensitive data, intellectual property and commercially sensitive data.

8.4 Life cycle safety of data

The life cycle of data refers to the entire process of data generation, data collection, data encryption or decryption, data standardization, data integration, data storage, data analysis, data sharing and disposal.

A holistic data security approach with appropriate and proportionate security measures should be introduced to deter and disrupt hostile, malicious, fraudulent and criminal behaviours or activities. The security approach should seek to preserve safety, authenticity, confidentiality, integrity, availability, provenance and reliability of integrated data, ensuring where possible that integrated data are free from danger or threat of unintended access and use in the whole life cycle of data.

All access to data should be tracked and monitored. The logging mechanism should be used in all environments during the data life cycle to track and analyse the possible abnormal conditions or errors. A data access monitoring mechanism should be established to link and implement automatic audit tracking the access to all system components with each user individually.

8.5 System security protection

When constructing new parameter data (referring to the exchange and sharing of data), a complete network isolation system should be ensured, i.e. a system with a lower security level is used for development and testing first, and then a new fully isolated high security is enabled when the test is completed.

Systemic security assessment should be established in which security controls, network connections and restrictions should be tested periodically, such as scanning vulnerability or testing penetration to ensure the system can identify and block unauthorized access.

It is recommended that a vulnerability management programme is created to upgrade scanning and anti-virus software or computer programs to ensure the system is not attacked.

Annex A
(informative)

Case studies

Table A.1 — Smart Park Project (Phase I, Window of Canton District B)

Project title	Window of Canton Smart Park Project (Phase I)
Project profile	Participants include construction developers (CCCC Fourth Shipping Engineering Bureau Co., Ltd.), ICT suppliers (iSoftStone Information Technology (Group) Co., Ltd.), and academia (Smart City Joint Laboratory, CSUS). The project aims to create a multi-in-one ecological, interactive and open new smart business space, and provide a more comfortable and convenient working and living environment for the enterprises and employees in the park and the management of the park. It uses the “digital twin” as the construction system and is based on the “BIM+IoT+AI” technology to realize the digitization of all elements of the park space equipment, visualization of the operation situation, centralized operation and maintenance management, scientific decision-making management and park resource sharing. For example, smart parking, visitor management, environmental monitoring, portal systems. The resources and channels integrated in the project process include: 1) Cloud data centre: This centre is a unified data centre for the project, providing physical space for the current and future data storage and processing of the Window of Canton Smart Park. 2) Network resource database: This database mainly collects data that provide high-speed network resources for park managers and users, including wired broadband networks, local basic networks, mobile and wireless broadband networks and internet of things data. 3) Data management database: This database mainly collects data related to the input system, information retrieval system, metadata management system, data display, operating system and other related management systems. 4) Smart parking database: This database is mainly used to automatically notify free parking space information, save parking time and input vehicle information at the same time, which can realize the reverse car search function and reduce the time to find a car. 5) Energy management database: This database is mainly used to collect real-time statistics of energy consumption in various regions, analyse the energy consumption status through the data collected in the database and assist managers in formulating energy-saving strategies. 6) Property management database: This database mainly collects and applies the online repair report function, automatically dispatches work orders, implements summary complaint and suggestion information and issues the corresponding department for processing.

Table A.1 *(continued)*

	7) Environmental monitoring database: This database mainly collects and realizes the monitoring of the main environmental parameters of the park, generates historical records of the park environment based on the data generated by the monitoring and assists in formulating the opening plan of the park's air-conditioning system and lighting system. 8) Face recognition database: This database mainly collects the relevant data for recognition by capturing facial pictures of people entering and exiting at the entrances and exits of the park and key places. Data can be integrated and applied to personnel and vehicle deployment control, employee attendance, access control systems, park consumption and smart office services. 9) Regional intrusion alarm database: This database mainly collects data generated by monitoring and fortifying key areas (e.g. the perimeter of the park, the roof). When someone breaks in illegally, the intrusion information is automatically notified to the property management personnel, thereby improving the security level of the park and reducing the occurrence of accidents. 10) Portal system database: This database mainly collects relevant data, including WEB portals and mobile portals (mobile service applications, WeChat official accounts). This project has just completed the first phase of construction. The research team is currently conducting a Phase II feasibility study. It is hoped that the success of this smart park platform can provide a reference for the construction of similar parks in China.	
Organization	CCCC Fourth Harbor Engineering Co., Ltd., ISoftStone Information Technology (Group) Co., Ltd, Smart City Joint Lab, CSUS	
Place	Haizhu District, Guangzhou City, Guangdong Province, China	
Time	June 2020 to October 2020	
Reference	None, the website is under construction	
Relevance to this document	4.2	Principles of heterogeneous data integration
	4.3	Data quality recommendations
	Clause 6	Data integration framework
	Clause 7	Data integration