Evaluation of Cities' Smartness through Multidimensional Platform of Performance Indicators

Dessislava Petrova-Antonova[1][a] and Martin Minkov[2]

[1]GATE Institute, Sofia University, 125 Tsarigradsko shose Blvd., Sofia, Bulgaria

[2]Faculty of Mathematics and Informatics, Sofia University, 5 James Bourchier Blvd., Sofia, Bulgaria

Keywords: Architecture of Platform, Conceptual Data Model, Performance Indicators, Smart City.

Abstract: A lot of cities are working on their digital transformation in order to deliver better living environment for their citizens. There are many research efforts focused on measuring the performance of such transformation through specific methodologies and indicators covering variety city dimensions. The complexity of cities as well as the heterogeneity of data that they produced bring challenges to development of platforms for multidimensional evaluation and smart level assessment of a city. In order to address such challenges, this paper proposes a conceptual data model and an architecture of a plat-form for performance assessment of smart cities. For assessment of economic performance, 4 indicators are considered to be implemented and are presented in the paper: Gross Domestic Product, New Business Registered, Median Disposable Income and Human Development Index. The proposed platform is designed to be integrated – continuously supplied with city data, scalable – open-ended for implementation of new indicators, multidimensional – designed to cover all city dimensions and agile – evolve in step with the changing requirements.

1 INTRODUCTION

A fast-growing percentage of the population lives in urban areas. The United Nations reported that 55% of the world's population lives in urban areas and is expected to increase to 68% by 2050 (UN, 2018). Realizing the trends in urbanization is a key factor for successful development. The pace of urbanization is projected to be the fastest in the low-income and lower-middle-income country. Cities face challenges in meeting the needs of their growing population, including but not limited to the energy, transportation, housing, healthcare and education. This requires informed decisions to be taken in a timely manner. City managers and communities are open to adopt new solutions that can bring more efficient management of resources, availability of relevant services, and long-term sustainable behaviour.

A lot of management systems have already incorporated the continuous inspec-tion of the effects of actions and process improvement. The management of the productivity through performance measurement receives a wider adoption. The city authorities and policy makers are aware with this idea, but its adoption brings a lot of challenges (Neumann et al., 2015):

- Monitoring and evaluation frameworks for cities have to be created;
- Lack of consistent policies implemented towards the smart city objectives;
- Vendor and technology locking of some solutions;
- Data privacy and security;
- Difficulties to identify solutions that offer benefits for all aspects of a city;
- Measuring "smart services" impact, performance and effectiveness.

It is acknowledged that adequate and sustained decisions need an accurate perception of the processes and environment. At the same time, examination of whether the expected effects match the actual results also need a way to explore the current situation, or at least some of its characteristics, and to assess the changes. However, the city is a complex system and capturing all its dimensions is a time and cost consuming process. Therefore, it is more effective to evaluate certain dimensions and give them quantified and qualitative

[a] https://orcid.org/0000-0002-9920-8877

Petrova-Antonova, D. and Minkov, M.
Evaluation of Cities' Smartness through Multidimensional Platform of Performance Indicators.
DOI: 10.5220/0010024401400146
In *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2020) - Volume 3: KMIS*, pages 140-146
ISBN: 978-989-758-474-9

expression through performance indicators. The indicators enable easier comprehension without losing representability.

In the context of smart cities, performance indicators are considered as a tool to check whether and to what extent the intended consequence of an action or policy is realized. There are two major groups of indicators – project indicators and city indicators. This work is focused on the latter, although many of the concepts might be valid for the former as well. Additionally, the indicators can be used not only for assessment of the city performance itself, but to compare the cities, especially when the effect of replicated solution should be measured. Thus, there are four cases of indicators' application: (1) assessment of performance of a single city in the present, (2) assessment of performance of a single city in the future, after some actions are completed or solutions are implemented, (3) comparison of the cities in the present, and (4) comparison of the cities in the future.

This paper proposes a *conceptual data model and an architecture of a platform* for flexible performance assessment of smart cities through a range of indicators covering all city dimensions such as living, people, transport, etc. The indicators give a valuable insight into the extent to which the city is becoming "smarter" and outline the driving factors for sustainable development.

The rest of the paper is structured as follows. Section 2 is devoted on the state of the art. Section 3 presents the conceptual data model, while Section 4 defines the indicators currently considered for implementation and validation of the platform. Section 5 describes the platform's architecture. Finally, Section 6 summarizes the paper and provides directions for future work.

2 STATE OF THE ART

An increasing number of smart city initiatives exist all over the world aiming to deliver better planned, more connected and more liveable cities. Amsterdam in the Netherlands, Barcelona in Spain and Stockholm in Sweden are remarkable examples for implementation of smart city vision. A significant number of events such as conferences and exhibitions dedicated on smart cities are held every year. In 2011, a global event Smart City Expo World Congress was launched in Barcelona (Smart City Expo World Congress, 2011). Annual country-specific events, such as, the Smart Cities Week in Washington DC (2018), the Telegraph Britain's Smart Cities Conference (2018) and Conference for South-East Europe in Sofia, Bulgaria (2018) are also organized.

There are a lot of FP7 and Horizon 2020 projects and research initiatives related to smart cities, such EIP-SCC Market Place, SMARTIE, EU CIP Open Cities, FIWARE, FINESCE, etc. All current activities are mainly focused on improving the current living conditions in cities and are often related to specific city dimensions such as e-ticketing, smart street lights, pollution reduction, etc. But what is beyond the smart city is the information-rich city presented with intelligent models that support planning, design and analysis of all city dimensions.

As was reported in our previous work, there are huge number of key performance indicators (KPIs) defined as well as several available tools and platforms to analyse and evaluate smart city's performance (Petrova-Antonova and Ilieva, 2018). Benchmarking procedures have been proposed for comparative analysis and ranking of cities (Garau et al., 2005; Giffinger, 2007, Shields and Langer, 2009; Afonso et al., 2015; O'Neill and MacHugh, 2015). On the other hand, there is a number of assessment procedures based of multidimensional approach for measuring effects of smart city initiatives (UN, 2007, Minx et al., 2010; Rosales, 2010; Priano and Guerra, 2014; Orlowski et al., 2016; Bosch et al., 2017; Tanda et al., 2017; Marijuan et al., 2017). To the best of our knowledge, most of proposed smart city evaluation approaches and instruments only allows to assess the current situation of a city without connecting it with a technological solution allowing for continuous monitoring and evaluation of city's digital transformation using urban data from stakeholders and physical objects.

3 CONCEPTUAL DATA MODEL

The analysis of a straightforward case of implementation – one which simply transmits already available values – is not applicable. Thus, let's examine a situation where data needs to be obtained from multiple sources in terms of different levels of distance from the primary data origin, as well as different level of transformation of the data. Furthermore, some sources may have already calculated indicators' values for other purposes and reduce the role of the platform as simple transmitter. Others might simply be a plain access point for raw sensor data and make the platform a primary processor. Finally, there are indicators' values that provide already pro-cessed data, which might serve a dual purpose – both as direct output and used in intermediary calculations within the platform (see Fig. 1).

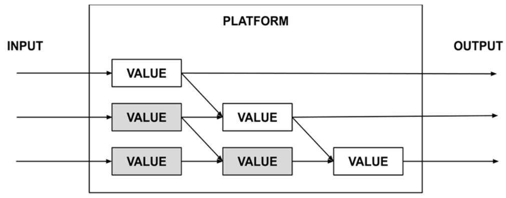

Figure 1: Different levels of processing within the platform.

Not all data incoming into the platform is raw data. There are two kinds of values calculated within the platform – intermediary ones and final indicators. And then, there those that have dual purpose – both as intermediary and final. Considering all of the above, it is proposed that data should be modelled within the platform in a general sense, suspending the existing distinctions between initial measurements, primary data, intermediary values and final indicator values. As a result, the conceptual model, shown in Fig. 2, is proposed.

The data becomes information only once it's been linked to a context (a meaning). For example, the percentage “56%” on its own does not indicate any usable resource. Attaching it to a label of "portion of the green areas that have tree plantations" al-lows us to know the fields to forests ratio. When the smallest piece of datum is considered it can be posed that this quantum of data cannot and should not be modelled for the purposes of smart city indicators. A first step is to combine it with a label (meaning) and end up with a "concept"–"value" pair. The “label” will be referred in the model as Value Concept (VC) – representing what are the semantics of the datum – while the datum itself is designated as a Resolved Value (RV). From this starting position there are four directions for further consideration:

- Situational reference – What the value is about?
- Resolution – How the value is arrived at?
- Origin – What is the lineage of the data that was transformed to the eventual resolved value?
- Classification of the concept with a scheme or framework.

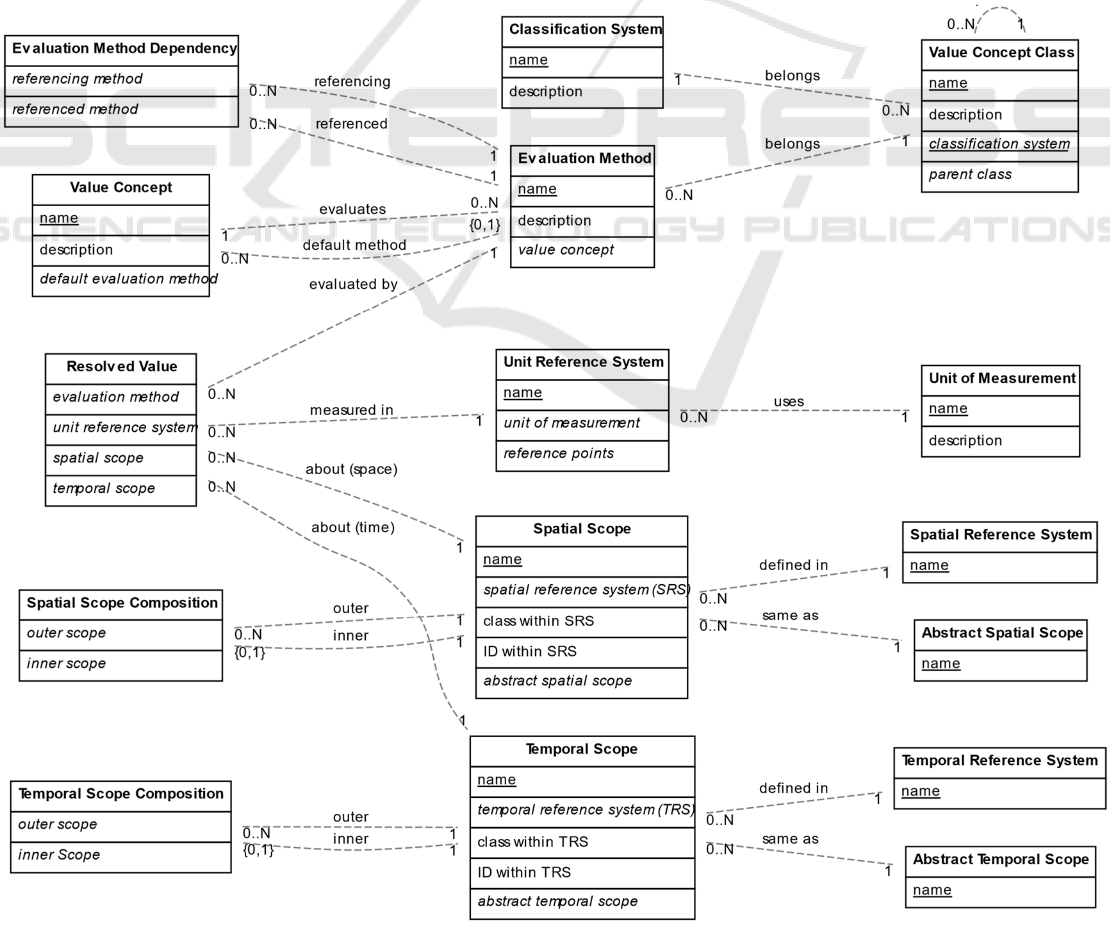

Figure 2: Conceptual Data Model.

On its own a named value remains sufficiently abstract to be actionable in practice as it lacks reference to the real world. An open question is what part of space and time does it apply to. Thus, the "concept" is what the value is, while the scope is where the value is valid. That is why a four-dimensional model is adopted, where a Spatial Scope (SS) and Temporal Scope (TS) are added to the value (Bosch et al., 2016).

There are several geographic reference systems, which not necessarily match or reference one another. A scope must be aware of which reference system it refers to, along with a specification of the object. Thus, a scope needs to refer to a scope reference system, a type within that reference system and an identifier (or specification) of an element of that type within that reference system. Due to possibility the same objects existing in different data sources to follow different reference systems, an Abstract Spatial Scope (ASS), which creates relations between them, is introduced.

The choice of unit reference system influences the compatibility of indicators. It is recommended that indicators should avoid being in absolute units and instead be either in ratios or be without a unit at all (Bosch et al., 2017). Suggested techniques for improving the comparability of values include normalization (mapping to a fixed scale, e.g., 0 to 10) and standardization, e.g., by using z-transformation (Giffinger et al., 2007). The Likert scale also allows for a unified representation of values, but the boundaries of the scale should be preliminary known. Therefore, the conceptual model requires an intermediary – Unit Reference System (URS) – which should specify both the unit as well as those boundaries as "reference points".

From the perspective of the platform's users when they request the value of a concept (for a specific spatial and temporal scopes) and receive it, they don't necessarily care whether it was calculated within the platform or it was taken from external source. When the indicator's value is calculated by the platform, a calculation method should be available. If the value is obtained form an external source, then it can be externally calculated or simply measured. The proposed conceptual model considers both calculation and measured (sensing) methods and describes them as an Evaluation Method (EM). Since a value concept might change the method for its evaluation while keeping the name, it is appropriate to link a resolved value with an evaluation method, and to define a second level reference to a value concept. Thus, the evaluation method reduces the value concept to an alias, a named pointer to a method or a resolved value's semantics. When the calculation of a value depends on other values, that relationship is represented in an Evaluation Method Dependency (EMD). The dependency complies the spatial and temporal scopes.

The general algorithm for resolution is presented in Fig. 3.

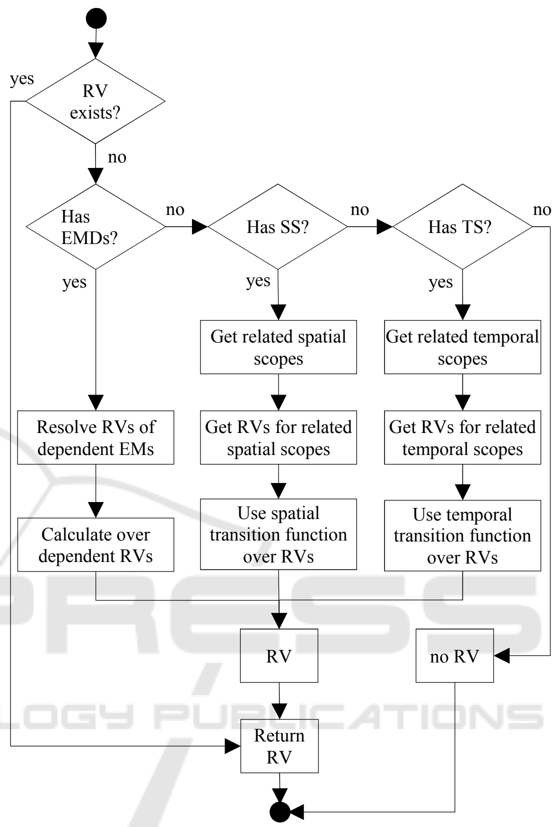

Figure 3: General Resolution Algorithm.

Another open question is the origin of the indicator's value. When the value is calculated within the platform the answer is straightforward. However, if a value is obtained from an external system, there are 2 possible situations: (1) if the value is directly obtained from the external system, then the latter can be seen as its origin and (2) if the value is calculated by the external system, but is obtained from a third-party system (e.g. by sensing), than the third-party system can be seen as its origin. Since the origin of a value does not affect the calculation process of the platform the origin of values will not influence the current design.

There are variety classification schemes that can be used for grouping of indicators and/or classifying them in hierarchical categories. For example, in our previous work the indicators are classified in six thematic areas, namely Smart Nature, Smart Living, Smart Mobility, Smart Governance, Smart People and Smart Economy (Petrova-Antonova and Ilieva, 2018). That is why the conceptual data model allows specification of indicator's classification schemes by introducing a Classification System. Using the Value Concept Class, a hierarchy of classes can be defined.

4 PERFORMANCE INDICATORS OF THE PLATFORM

The current implementation of the platform considers 3 indicators for economic performance, described in CITYkeys project (Bosch, et al., 2017) and Human Development Index, proposed by the United Nations Development Programme (UN, 2019):

- Gross Domestic Product (GDP) – gross domestic product per capita;
- New Business Registered (NBR) – number of new businesses per 100,000 population;
- Median Disposable Income (MDI) – median disposable annual household income;
- Human Development Index (HDI) – assessment of average achievement in the most important dimensions of human development, measured by 3 separate indexes: Life Expectancy Index (LEI), Education Index (EI) and Gross National Income Index (II), based on GNI Index.

The GDP is a widely accepted measure for economic performance, which provides an aggregate measure of production. The indicator is calculated according to the following equation:

$$GDP_{per\ capita}=\frac{GDP}{population} \quad (1)$$

The NBR assesses the overall business climate and entrepreneurship attitude. It is calculated according the following equation:

$$NBR_{per\ 100000\ capita}=\frac{NBR}{population\times 100000} \quad (2)$$

The MDI is related to economic wealth related to improve access to quality education, housing and healthcare. The total disposable household income is computed as total household gross income, reduced to regular taxes on wealth, regular inter-household cash transfer paid, tax on income and social insurance contributions (Eurostat, 2011). The median is the middle value, i.e. 50% of all observations are below the median value and 50% above it (Bosch et al., 2017), and is calculated as follows:

$$\mathrm{MDI}_{\mathrm{household}}=\frac{\mathrm{Income}}{\mathrm{household}} \quad (3)$$

The HDI is a geometric mean of normalized indices for each of the three dimensions of human development, described so far, as follows:

$$HDI=\sqrt[3]{LEI\times EI\times II} \quad (4)$$

The life expectancy at birth is defined as "the number of years a new-born infant could expect to live if prevailing patterns of age-specific mortality rates at the time of birth were to stay the same throughout the child's life" (UN, 2010). The LEI is based on a minimum value of 20 years and a maximum value of 85 years. The II is calculated using the natural logarithm of GNI per capita adjusted by PPP, which minimum value is \$100 and maximum value is \$75,000 (UN, 2015). The EI is composed by two indicators, namely the Mean Years of Schooling (MYS) for adults aged 25 years and older, and the Expected Years of Schooling for children of school entering age. It is defined as follows:

$$EY=\frac{MYS+EYS}{2} \quad (5)$$

The LEI, II, MYS and EYS indexes are calculated as follows:

$$Index=\frac{actual\ value\text{-}minimum\ value}{maximum\ value=minimum\ value} \quad (6)$$

5 PLATFORM ARCHITECTURE

The high-level architecture of the platform is shown in Fig. 4. The Data Store is implemented as a relational database. Each concept from the conceptual model has a corresponding table in the database. The data can be collected automatically via APIs or manually imported by the users. The automatic data acquisition supports both pull and push methods. The push method requires the data to be transmitted in real-time and typically it belongs to a single primary source. The pull method allows data to be batch processed at different intervals (e.g. via a schedule). It supports cross system integration, since data can be integrated from many sources. Thus, the automatic data acquisition supports both batch and real-time dataflows through pull and push APIs.

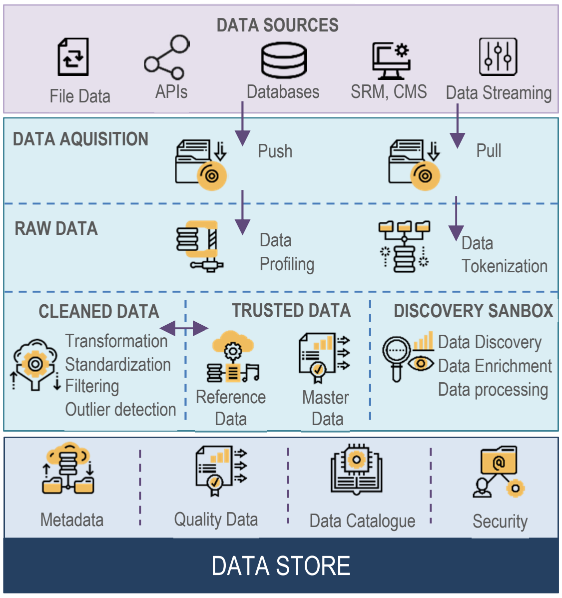

Figure 4: Architecture of the Platform.

One of main features of the platform is data collection from multiple sources. Since the raw data cannot be used directly for calculation of the indicators, different techniques should be applied to provide a semantic interoperability, as follows:

- Data profiling – process raw data to collect statistics and define rules and constrains;
- Data tokenization – replace sensitive data with tokens (random strings of characters) that keep the essential information about the data without compromising security.
- Data filtering – remove records, which are not compliant with data rules and constraints (duplicate rows, incorrect information, etc.);
- Data transformation – transform data according to the rules and constrains, including normalization of values;
- Data standardization – convert the structure of a dataset into a uniform format;
- Outliers detection – find outliers with extreme values that deviate from other observations on data.

Data enrichment is an additional step towards producing high quality datasets. The data from a given dataset are merged with third-party data from external authoritative source to produce more deep insight. Along with data quality, the management of metadata are the second important feature of the Data Store. The metadata make it easy to find and process particular instances of data. In order to increase the discoverability of datasets and data services the Data Store relies on the Data Cata-log Vocabulary (DCAT) proposed by W3C. DCAT enables datasets and data services to be described in a catalogue using a standard model and vocabulary facilitating the consumption and aggregation of metadata (W3C, 2020). The data quality and metadata are closely connected. The metadata put the data in a context, and thus turn facts into actionable information. Both data and metadata are parts of the data catalogue, which acts as an inventory of data assets in the Data Store. In addition, it provides secure access to the data assets based on preliminary defined policies.

6 CONCLUSIONS

The cities are centres for people and economic activities and therefore the main drivers of the sustainable and inclusive growth. According to the globally accepted United Nations Sustainable Development Goals, they play a crucial role well-being of the citizens by providing an access to the employment, health and educations opportunities as well as to civic and social engagement. The greater opportunities in cities bring in turn a greater risk. That is why their performance should be continuously monitored and assessed based on clearly defined indicators. For example, differences in GDP growth by distance to large cities provide insight about their impact on the economy. The rising of the population density requires new solutions to be found. In such context, a platform for flexible performance assessment of smart cities is proposed. It is able to handle a range of indicators covering all city aspects such as living, people, transport, etc. The indicators give a valuable insight into the extent to which the city is becoming "smarter" and outline the driving factors for sustainable development. The conceptual data model of the platform and its architecture are presented. Sample indicators for economic performance that are considered to be implemented in the platform are also described.

ACKNOWLEDGEMENTS

This research work has been supported by GATE project, funded by the Horizon 2020 Widespread-2018-2020 Teaming Phase 2 programme under grant agreement no. 857155 and Big4Smart project, funded by the Bulgarian national science fund, under agreement no. DN12/9.

REFERENCES

Afonso R., dos Santos Brito K., Alvaro A., 2015. Brazilian Smart Cities: Using a Maturity Model to Measure and Compare Inequality in Cities. 16th Annual International Conference on Digital Government Research, pp. 231-238.

Bosch P., Jongeneel S., Rovers V., Neumann H., Vielguth S., Huovila A., Seppä I. P., Penttinen T., van der Heijden R., Kotakorpi E., Pangerl E., Neralić A. M., Funes D. S., 2016. Definition of data sets. CITYKeys project 2.1 Deliverable.

Bosch P., Jongeneel S., Rovers V., Neumann H., Airaksinen M., Huovila A., 2017. CITYkeys list of city indicators. CITYKeys project.

Bosch P., Jongeneel S., Rovers V., Neumann H., Airaksinen M., Huovila A., 2017. CITYkeys indicators for smart city projects and smart cities: Deliverable D1.4.

Eurostat, 2011. EU-SILC Description Target Variables, https://ec.europa.eu/eurostat/documents/1012329/6070906/Household+data+-+income.pdf.

Exhibition & Conference for South-East Europe in Sofia, 2018, https://www.clustercollaboration.eu/profile-events/6th-exhibition-conference-south-east-europe-sofia-bulgaria.

Garau Ch., Masala F., Pinna F., 2015. Benchmarking Smart Urban Mobility: A Study on Italian Cities. ICCSA 2015, Part II, LNCS 9156, pp. 612-623.

Giffinger R., Fertner Ch., Kramar H., Kalasek R., Pichler-Milanović N., Meijers E., 2007. Ranking of European medium-sized cities. Centre of Regional Science, Vienna University of Technology.

Marijuan A., Etminan G., Möller S., 2017. Smart cities information system: Key performance indicator guide.

Minx J., Creutzig F., Medinger V., Ziegler T, Owen A., Baiocchi G., 2010. Development a pragmatic approach to assess urban metabolism in Europe. European environment agency.

Neumann H. et al., 2015. Overview of the Current State of the Art. CITYKeys project 1.2 Deliverable.

O'Neill K. MacHugh I., 2015. Urban environment good practice & benchmarking report. RPS Group, Ireland.

Orlowski C., Ziolkowski1 A., Orlowski A., 2016. High-Level Model for the Design of KPIs for Smart Cities Systems. TCCI XXV, LNCS 9990, pp. 1-14.

Petrova-Antonova D., Ilieva S., 2018. Smart cities evaluation – A survey of performance and sustainability indicators. 44th EUROMICRO Conference on Software Engineering and Advanced Applications (SEAA), Prague, Czech Republic, pp.486-493.

Priano F., Guerra C., 2014. A framework for measuring smart cities. 15th Int. Conf. on Digital Government Research, pp. 44-54.

Rosales N., 2010. Towards a design of sustainable cities: incorporating sustainability indicators in urban planning. 46th ISOCARP Congress, pp. 1-11.

Shields K., Langer H., 2009. European Green City Index. Siemens AG.

Smart Cities Week in Washington DC, 2018. https://www.smartcitiesweek.com/2018-washington/.

Telegraph Britain's Smart Cities Conference, 2018.https://www.telegraph.co.uk/business/britains-smart-cities/.

Smart City Expo World Congress, 2011. http://www.smartcityexpo.com/en/the-event/past-editions-2011.

Tanda A., De Marco A., Rosso M., 2017. Evaluating the Impact of Smart City Initiatives. In: 6th International Conference on Smart Cities and Green ICT Systems, pp. 281-28.

United Nations, 2019. Human Development Index, http://hdr.undp.org/en/content/human-development-index-hdi.

United Nations, 2007. Indicators of Sustainable Development: Guidelines and Methodologies., 3rd Edition.

United Nations, 2018. Revision of World Urbanization Prospects, https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html.

United Nations, 2010. The Real Wealth of Nations: Pathways to Human Development, Human Development Report, http://hdr.undp.org/sites/default/files/reports/270/hdr_2010_en_complete_reprint.pdf.

United Nations, 2015. Training Material for Producing National Human Development Reports, http://hdr.undp.org/sites/default/files/hdi_training.pdf.

W3C, 2020. Data Catalog Vocabulary (DCAT) - Version 2, https://www.w3.org/TR/vocab-dcat/.