CITIES
Centre for IT Intelligent Energy Systems

🔗 smart-cities-centre.org   🐦 @CITIES_Centre   in cities-innovation-center

DTU
AALBORG UNIVERSITY DENMARK

/nnovation Fund Denmark

CITIESData
a smart city data management framework

Xiufeng Liu[1], Alfred Heller[2], Per Sieverts Nielsen[1]
1 DTU, Department Management; 2 Niras

Introduction

Modern cities which have a wide use of information and communication technologies are coined as smart cities. Digitalized smart cities produce various type of data, which come from heterogeneous sources including various types of the Internet of Things such as traffic, weather, pollution, noise, and portable devices. The continuously increasing production of data has also raised focus on privacy protection.
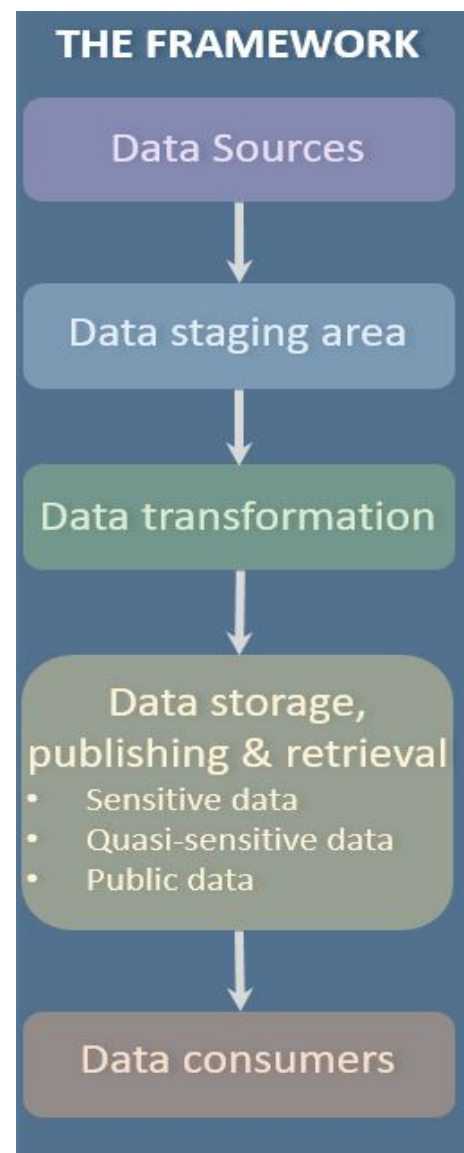
Therefore, CITIESdata platform focuses on not only management of data publishing services, but is also designed to streamline the whole data management process in a secure environment, including collecting, cleansing, anonymization, and data publishing. In particular, data quality issues, privacy protection, and data exchange are emphasized.

Types of data

.Sensitive: Data have the attributes that can directly identify an individual (e.g., social security number., bank account, etc.)

.Quasi-sensitive: Data have the attributes that can identify an individual when linking to the external sources (e.g., age, gender, address, etc.)

.Open/Public: Data are published or shared on an open data platform

THE FRAMEWORK

Data Sources

Data staging area

Data transformation

Data storage, publishing & retrieval
- Sensitive data
- Quasi-sensitive data
- Public data

Data consumers

Data preprocessing

In the CITIESData framework, the goal is to use the data for analytics where data quality is critically important to ensure the correctness of data models and analytic results. Furthermore, smart city data may contain sensitive information that requires desensitization to protect privacy. In this way three functional modules for pre-processing the data were developed.

.Data cleansing: Since poor data quality can lead to unreliable results of any data analytics, a data cleansing process is required. In the suggested framework, an automatic data cleansing can be done by using rule-based methods, or by a manually visible analytics approach.

.Data quality checking: It is a good practice to check the data quality before and after the data cleansing, because the results will provide an indication of data quality and give a reference to users before the data are used. There are two broad data quality checking methods identified, data profiling and data mining.

.Data anonymization: As some data may contain sensitive information, they have to be desensitized through a sequence of anonymization operations. In other words, these are the approaches that seek to hide the identity and the sensitive data of owners.

REMARK

There are various use cases for handling smart city data, which is why it is challenging to use a uniform data processing system or platform.

A real case-test was used, in which energy data sets were used to evaluate the CITIESData platform and validated its effectiveness and efficiency. The proposed framework supports several underlying data processing systems. This feature makes the framework distinct to others, because it supports processing different types of smart city data from extracting to publishing within the same platform.

We believe that the proposed framework is a good candidate for a common solution for smart city data management.

About the authors

Xiufeng Liu (xiuli@dtu.dk), Senior Researcher, Department of Management Engineering at DTU, Technical University of Denmark, Kgs. Lyngby, Denmark

Alfred Heller (ahr@niras.dk), Associate Professor, Department of Civil Engineering at DTU and now works for NIRAS A/S, Sortemosevej 19, 3450 Allerød, Lillerød, Denmark

Per Sieverts Nielsen (pernn@dtu.dk), Senior Researcher, Department of Management Engineering at DTU, Technical University of Denmark, Kgs. Lyngby, Denmark