



Informatics Department
Aristotle University of
Thessaloniki



Business
Administration
TEI of Thessaly

Smart Cities as Data Engines

Athena Vakali, Vaia Moustaka, Leonidas G. Anthopoulos



Thessaloniki, May 2018

Smart Cities Fundamentals



Smart cities are highly sophisticated systems that attract

- ✘ **governments, policy makers and municipalities** (e.g., EU smart cities initiative, World Smart City Forum, Smart City Business Institute)
 - have structured progressive policies to deal with issues like **urbanism** and climate change, United Nations 2030 Agenda for sustainable development
- ✘ **industries with the leading role of the Information & Communication Technologies** (e.g., CISCO, IBM, Libelium, Ericsson, etc.)
 - define a new competitive market that is estimated to become **dominant by 2030**
- ✘ **Scientists**
 - investigate the future of an interdisciplinary and much promising domain that combines studies like engineering, ICT, humanities, ethics, political science, etc.

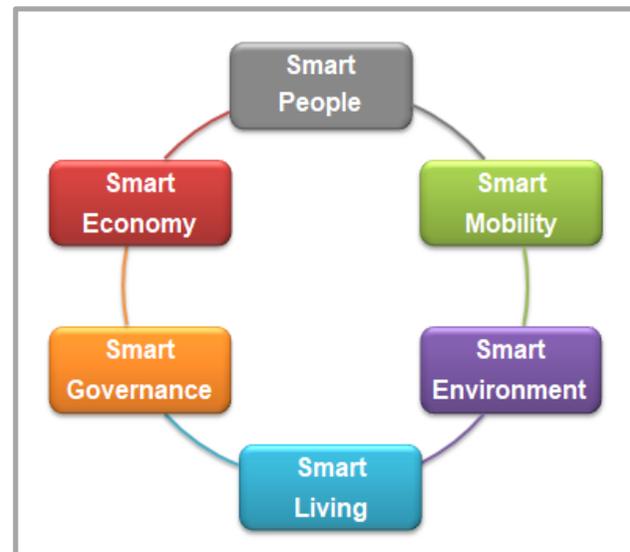
Definition (ISO/IEC, 2015)

*“Smart city is an innovative city that **uses ICT** and other means to improve quality of life, efficiency of urban operation and services, and competitiveness, while ensuring that it meets the needs of present and future generations with respect to economic, social and environmental aspects”*

Smart Cities Fundamentals

Smart Cities Dimensions

The conceptual model of Giffinger & Gudrun (2011) recognize **6 smart dimensions**



Smart Cities Performance

The performance of smart dimensions and their sub-dimensions are evaluated by **Key Performance Indicators (KPIs)** adopted by **ISO/TR 37150** (ISO/IEC, 2015)

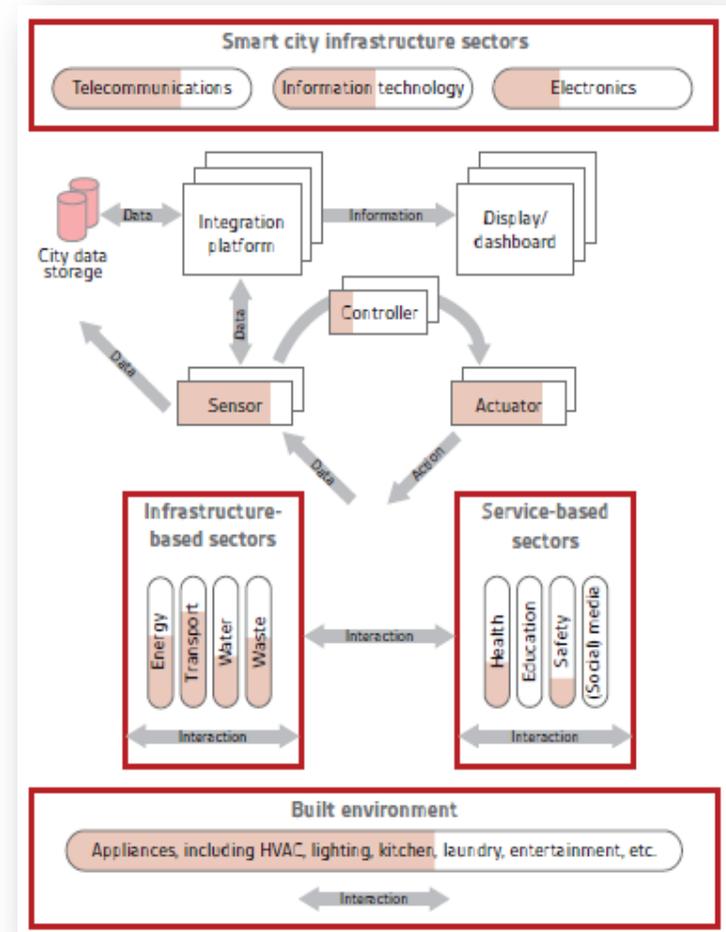
Smart Cities Fundamentals



Smart City as a “data system”

British Standard Institution (BSI) views SC as a **system that consists of several subsystems** (so-called “infrastructure-based” and “service-based” sectors) where **data** “

1. is **produced** and collected via sensors from different hard facilities (energy, transport, etc.) or in service-based sectors (health, safety and social media)
2. is **stored** in SC data storage and other infrastructures (also for telecommunications & electronics)
3. **Analyzed and visualized** on city dashboards or delivered to services’ end-users

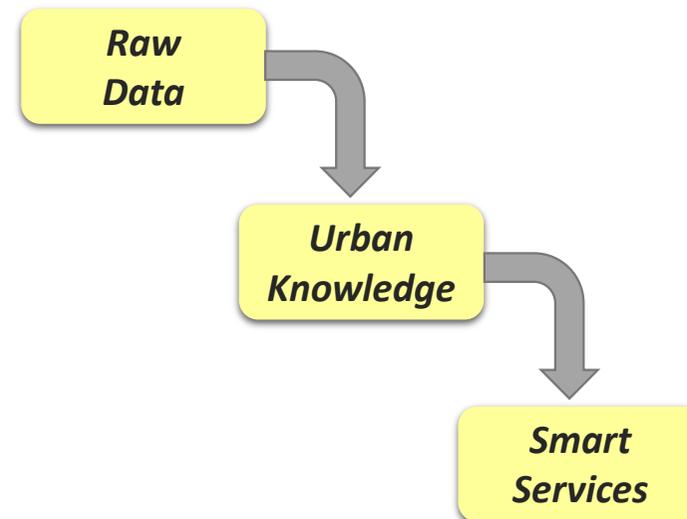
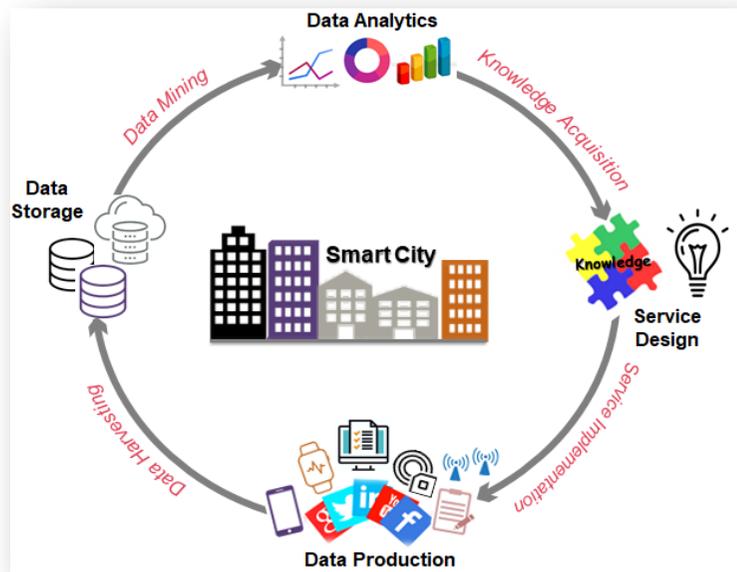


Smart Cities as Data Engines

Smart City as a “data engine”

Based on BSI’s approach we consider that *smart city behaves as a “data engine”, which constantly produces and consumes data*

- ✘ data flows follow a circular process, since –even during the last step– the analyzed data is stored and compared with other collected information or it returns back to the community as the context of new smart services



Smart Cities as Data Engines



Urban Data Sources

Large scale heterogeneous datasets are produced by:

- ✘ **fixed or portable sensing devices** (e.g., environmental sensors, traffic sensors, motion detectors, mobile devices, wearable devices, etc.)
- ✘ **web or mobile applications** (e.g., social networks, smart applications, etc.)

Urban Data Production

- ✘ **Internet of Things (IoT) data production** from sensors and actuators embedded in physical objects which are linked through wired and wireless networks - This “umbrella” term involves all the interconnected smart devices, such as RFID tags, sensors, cameras, mobile devices, etc.
- ✘ **Crowd-sensing data production** coming from the engagement of a defined “crowd” of individuals for obtaining required services, contents or ideas.



Smart Cities as Data Engines



Data Harvesting

“Data Harvesting is the gathering of data from numerous disparate databases into a single database from which it can be re-published in a unified manner”

Data Pre-processing

Urban *raw data* is characterized by heterogeneity



Data Pre-processing process is necessary as it affects the quality of collected data

- *Data Cleaning*
- *Data Integration*
- *Data Transformation*
- *Data Discretization*



Data Storage

- ✘ Depending on the *portability and usability requirements*, the pre-processed data sets are stored either in *traditional databases (DBMS)* or in *cloud storages*
- ✘ According to *data type*, data sets are stored either in *Graph DBMS*, or in *DBMS/SQL*, or in *NoSQL* (key-value stores, document stores, column-family stores, graph databases), or in *other data storages*

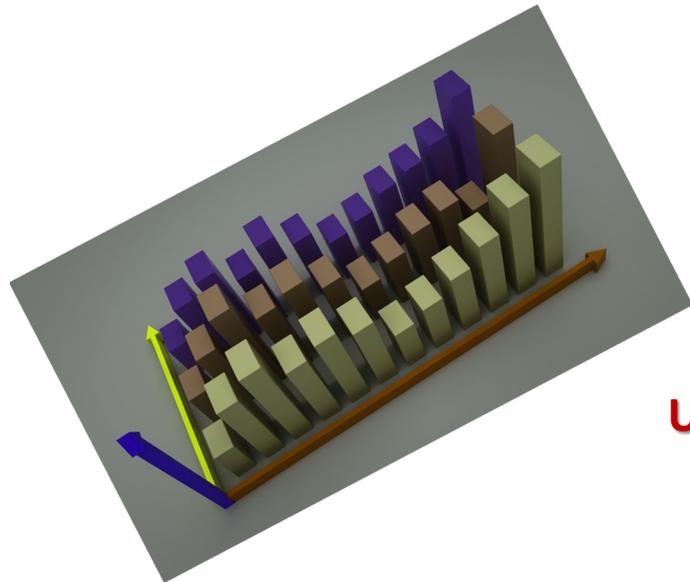


Smart Cities as Data Engines



Data Mining (DM) *a process of discovering various models, summaries and derived values from a given collection of data*

- ✘ **Descriptive** and **Predictive methods** are used for i) searching **non-trivial** information and patterns, and ii) predicting unknown values
- ✘ popular methods involve: (i) Clustering, (ii) Classification, (iii) Regression, (iv) Summarization, (v) Dependency Modeling, etc
- ✘ DM incorporates and utilizes many techniques and methods from other fields such as *statistics, machine learning, visualization, fuzzy logic, etc*



Urban Data Analytics

- ✘ Encapsulate techniques that are used to analyze and acquire profound knowledge out of urban data such as data mining, machine learning, statistical/predictive/graph analysis, etc.
- ✘ The gained **insights** by data analytics, can synthesize the **city profile** bringing out the urban potentials but also city's problematic issues

A systematic review for smart city data analytics: Methodology & Findings

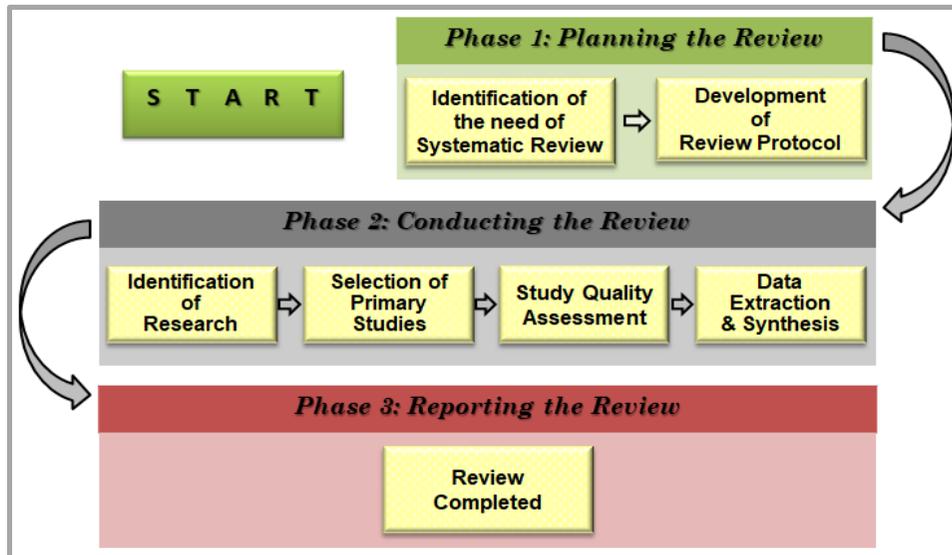


Aim

To connect the pieces between Data Science and SC or explicitly investigating data harvesting and data mining under the SC lens, while no similar work could be located.

Methodology

Initial search for relevant studies on the Internet for the period 1996–2017, returned 2,312 articles → **Kitchenham’s Guidelines**



Stage 1.1: Identification of the need of Syst. Review → Aim

Stage 1.2: Development of Review Protocol → Analysis & description of the actions that have to take place before the Conducting Phase

A systematic review for smart city data analytics: Methodology & Findings



Stage 2.1: Identification of Research → Res. Questions Definition

- RQ1. How many research studies exist addressing data harvesting and data mining processes in SC?
- RQ2. Which **methods** were used for the harvesting and mining of urban data?
- RQ3. Which smart **services** utilize urban data in smart cities?
- RQ4. What are the most common **sources** and types of storage of urban data?
- RQ5. Which smart **applications** utilize or produce urban data?

Stage 2.2: Selection of Primary Studies

- ✘ **Resources:** Google Scholar, Scopus, IEEE Xplore , and Science Direct
- ✘ **Search Terms:** «“Data Harvesting” AND “Smart Cities”» and «“Data Mining” AND “Smart Cities”»
- ✘ The search has been conducted for the **period 1996–2017** and was based on the *title*, the *keywords* and the *citations of the articles* to get the most relevant articles as search results

A systematic review for smart city data analytics: Methodology & Findings



✘ ***Inclusion/exclusion selection criteria*** according to them the candidate articles *are* evaluated and the final sample of the included articles is determined.

☐ Inclusion Criteria

- Articles published in Journal / Conference in which correspond at least 3 articles from those have found during the selection. This criterion applies only to the search string «“Data Mining” AND “Smart Cities”»
- Articles that perform at least 1 study that analyzes the data harvesting processes/data mining processes on smart cities

☐ Exclusion Criteria

- Articles performing studies related to smart services and not to smart cities
- Articles performing studies referred only to data harvesting/data mining processes or only to smart cities

Stage 2.3: Study Quality Assessment → Criteria to make certain the high quality level of the included articles

- ☐ the description of the data
- ☐ the availability of the data
- ☐ the description of the used methodology
- ☐ the presentation of the results

A systematic review for smart city data analytics: Methodology & Findings

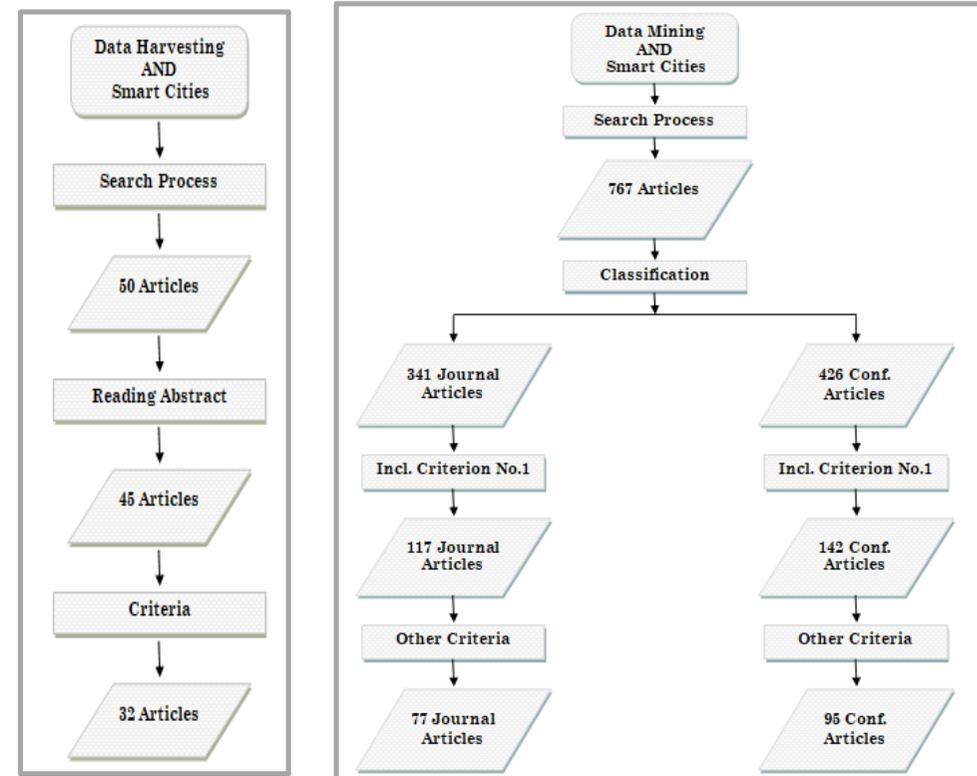
Stage 2.4: Data Extraction & Synthesis

Data features extracted from each article:

- i. Authors, publication source & year of publication (RQ1)
- ii. Type of article (Journal/Conf) (RQ1)
- iii. Data harvesting & analysis methods (RQ2)
- iv. Smart city dimensions & smart city services (RQ3)
- v. Urban data sources & urban data types (RQ4)
- vi. Smart applications (RQ5)

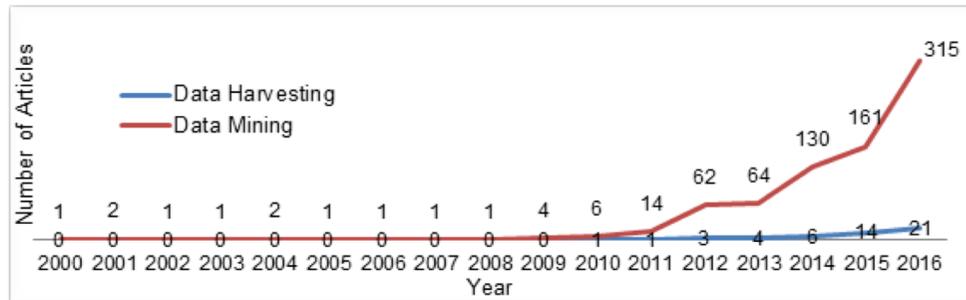
817
Articles → **204**
Articles

Selection process of the articles final sample (1996-2016)

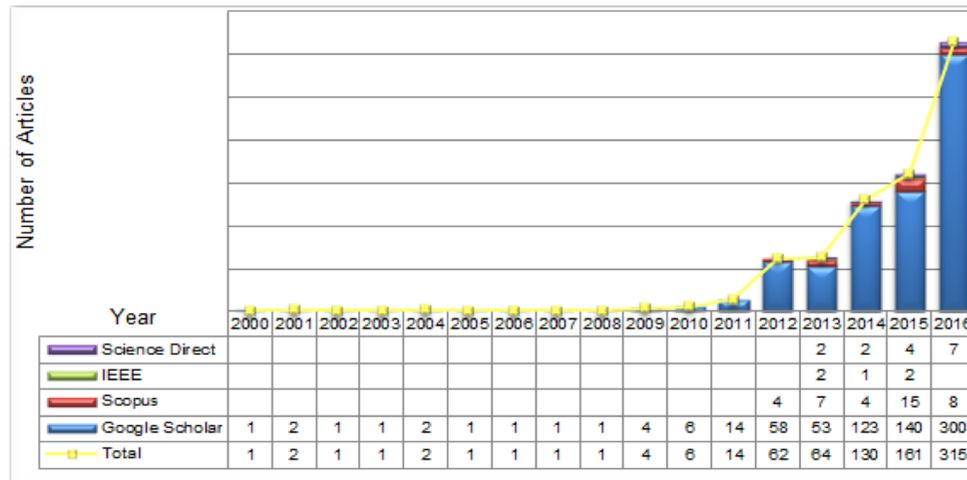


A systematic review for smart city data analytics: Methodology & Findings

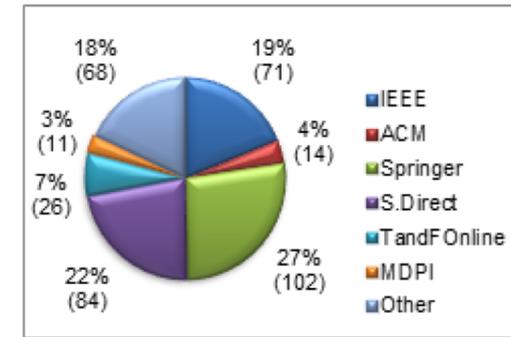
Answer to RQ1



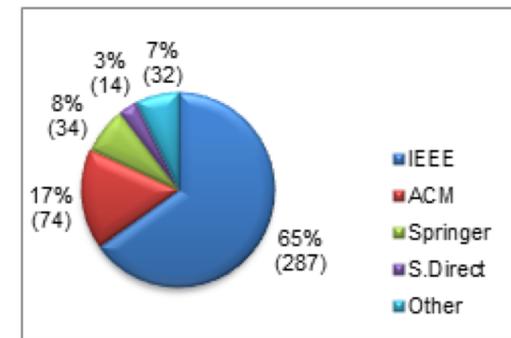
Rate of published articles (yearly)



Number of published articles for «“Data Mining” AND “Smart Cities”» search term /digital library / year



(a) Journal articles



(b) Conference articles

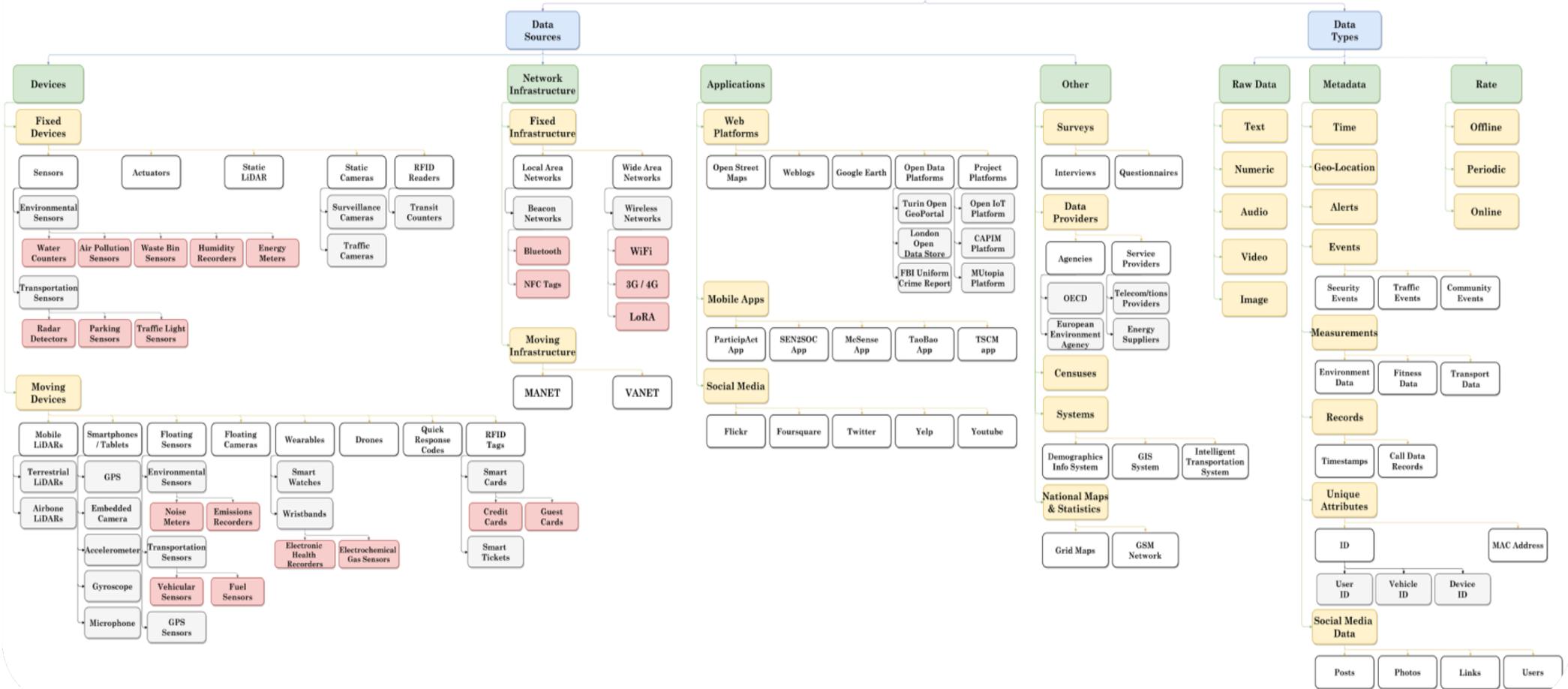
Articles per publication source
in period 1996-2016

DMS Taxonomy

Answers to RQ4 & RQ5



D TAXONOMY Data Production

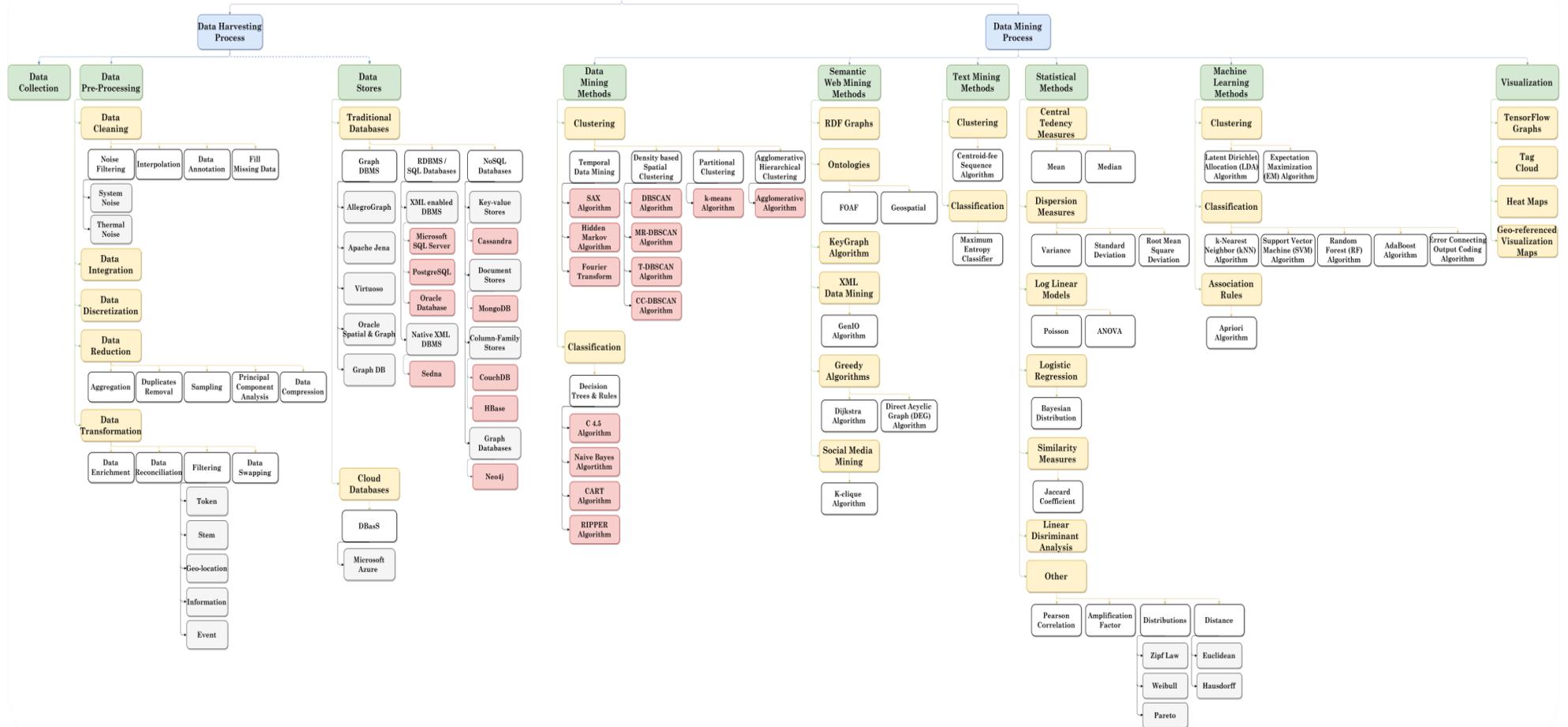


DMS Taxonomy



Answer to RQ2

M TAXONOMY Methods

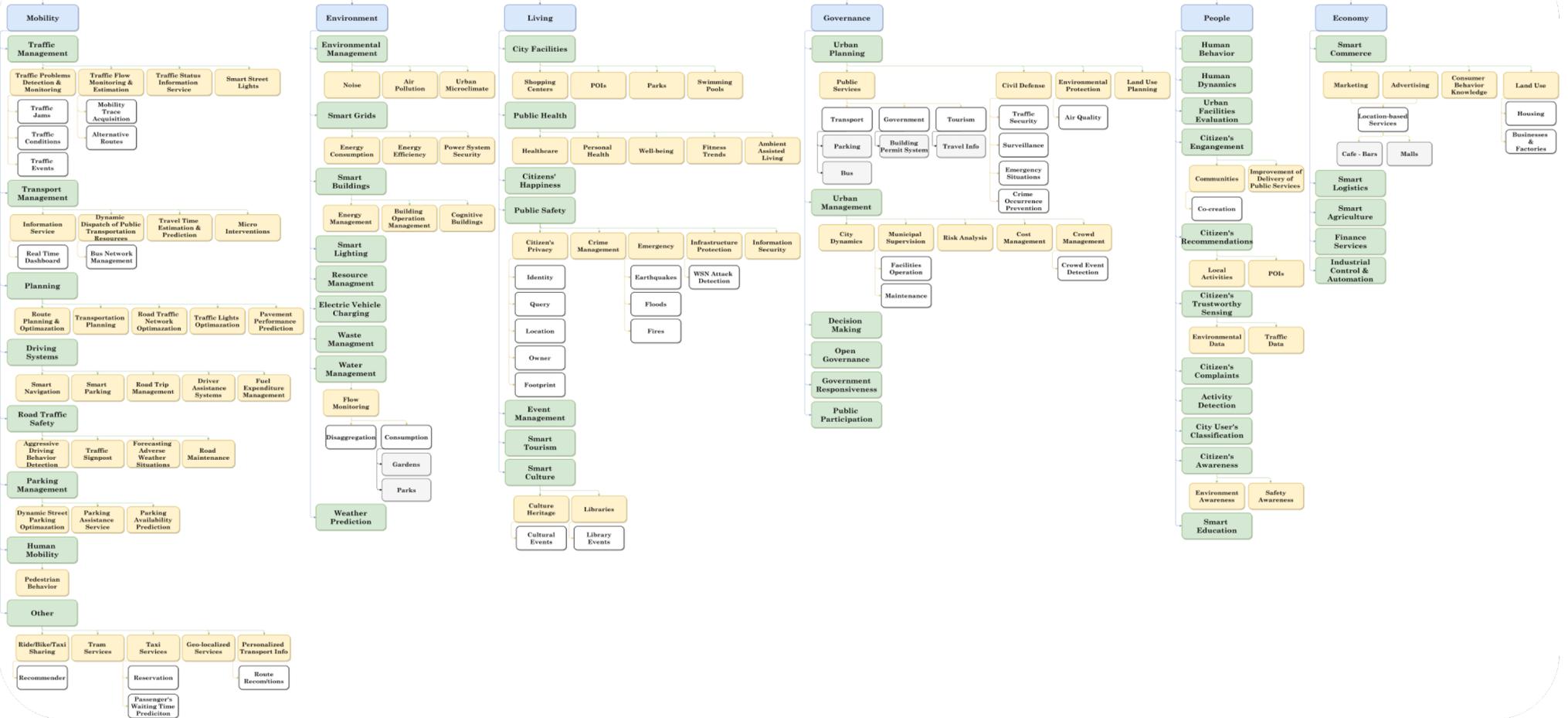


DMS Taxonomy

Answer to RQ3



S TAXONOMY Smart Services



Vision & Impact

Ongoing State-of-the-art & Trends (2017)

Resource	“data harvesting” AND “smart city”		“data mining” AND “smart city”	
	Initial results	Articles after screening	Initial results	Articles after screening
Google Scholar	44	24	1,330	548
Scopus	1	1	73	69
IEEE Xplore	0	0	27	10
Science Direct	4	4	17	11

✘ Data Production

- IoT & Crowd-sensing
- The results adjust to the generated DMS Taxonomy

✘ Data exploitation

- Similar data collection methods, storage resources & analysis techniques / algorithms
→ validation of the accuracy of the identified DMS taxonomy
- Big data & open data still attract scientific attention

Vision & Impact



× New types of smart services

- ❑ *smart food* (smart living)
- ❑ *transportation resilience* (smart mobility)
- ❑ *energy usage patterns for load prediction* (smart environment & smart governance)
- ❑ *indoor space quality* it is related with smart buildings and it is measured by human behavior
- ❑ *crime prediction* via criminal behavioral analysis (smart living)

...

× New trends ...

- ❑ an **increasing shift from SC smart dimensions to SC smart services** : SC architectural dimensions are surpassed by smart services (health, food, traffic, buildings, waste management etc.) advanced by DH and DM techniques
- ❑ **Emerging topics appear regarding user behavioral analysis & cyber-physical systems analysis**

Vision & Impact

- ✘ According to Syst. Review, most case studies are located in **Europe**, while North America and Asia follows
 - Validation of the continuous political support that smart cities gain momentum in Europe, which can be justified by **corresponding policies & funding opportunities** (i.e., Horizon 2020, URBACT, etc.)



(Source: Siemens, 2017)

Thank you for your attention!

