Preparing Smart Statistics

Istat's experience

Massimo De Cubellis - ISTAT

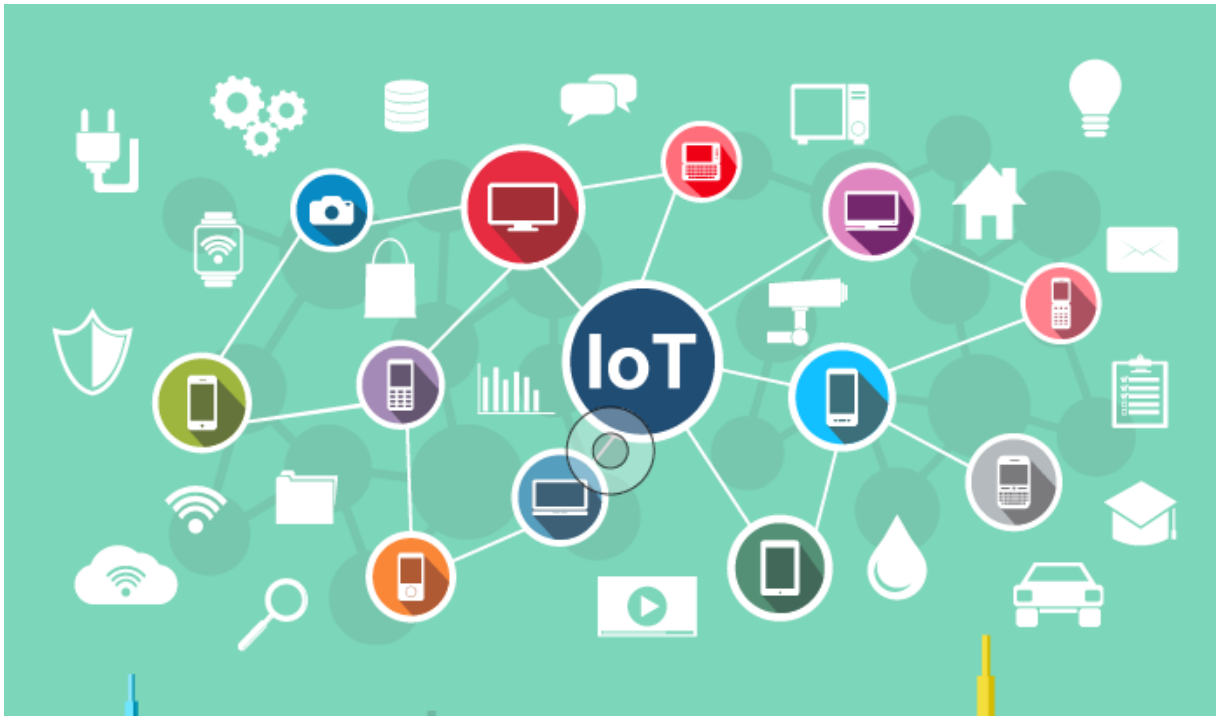
Index

1. IoT and Smart Statistics
2. Trusted Smart Statistics (TSS)
3. The role of Istat and others European National Statistical Institutes (NSIs)
4. The project ESSNet Big Data II / The WP-L
5. Smart Cities
    1. Case study - H2020 Lighthouse projects
    2. Case study - Air pollution
6. Smart devices

IoT and Smart Statistics

IoT and Smart Statistics

IoT (Internet of Things) is a neologism related to the extension of Internet to the world of objects (*Wikipedia*)



IoT and Smart Statistics

Objects (or things) mean the devices used in various contexts (urban, domestic, personal, industrial, etc.) that are the basis of the Internet of Things (IoT).

Their main features are:

- identification
- connection
- location
- to be able to process data
- to be able to interact with the external environment

Let's think to the smartphones …

- identification (IMEI code)
- connected at network (4G/5G connections)
- location (GPS)
- to be able to process data (Apps)
- to be bale to interact (interaction with other devices such as smart watch, smart TV, ecc.)

IoT and Smart Statistics

In this context, **smart** means: technologies incorporated in smart systems that make it possible to transform data into statistical information

Smart Statistics: statistics produced using data from smart devices

It means:

- statistics produced **for external purposes** to the functionality of the device, using the data that the smart devices collect (i.e. : we could use the data collected by the GPS of the smartphone to produce statistics on how the people move in their city)

- statistics produced autonomously by the device or by the smart system **to perform own functions** (i.e.: a system of smart traffic lights for traffic management uses the historical series on daily / hourly traffic to schedule the time of traffic lights on / off)

Trusted Smart Statistics (TSS)

What does *Trust* mean ?

Trust in input data

- data have to be veracious
- data provision have to be continuos and stable (quality)

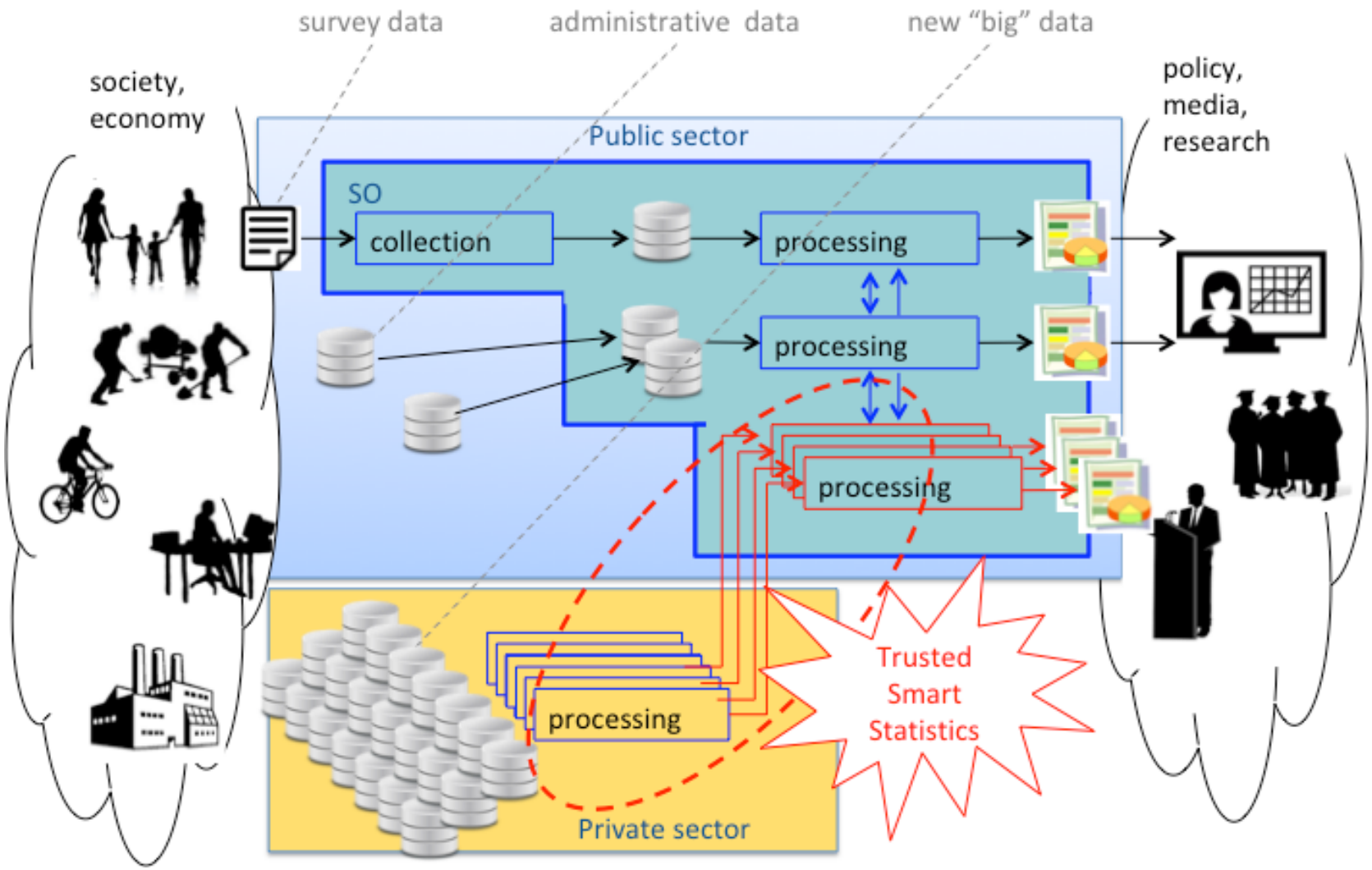
Trust in processing

- Data have to be processed only:
  - for agreed purposes
  - with agreed and transparent methods

Trust in Output

- statistics have provide quality information
- the output have to be in compliance with legal provisions
- privacy & confidentiality have to be guaranteed
- statistics have to correspond to the perceived situation

A reference architecture for TSS



Reference: «Towards a Reference Architecture for Trusted Smart Statistics" - ESS Task Force on Big Data and Official Statistics (Fabio Ricciato, Michail Skaliotis, Albrecht Wirthmann, Kostas Giannakouris, Fernando Reis)

Istat

The role of National Statistical Institutes (NSIs)

The role of NISs (European level)

Scheveningen Memorandum (February 2014)

- Eurostat encouraged the experimentation of new techniques that use Big Data for the production of official statistics
- It marked the start of exploring big data for use in official statistics within the European Statistical System (ESS)

Bucharest Memorandum (October 2018)

- the *DGINS* (Directors Generals of the National Statistical Institutes) recognized that the fast developments in the IoT, and the penetration of digital technologies into everyday life represent:
  - challenges and opportunities
  - threats and responsibilities
- It acknowledges the achievements reached so far (with projects *ESSNet Big Data I*) and identifies priorities to further develop the European Statistical System in a datafied society towards TSS

The projects ESSNet Big Data

Objective: to integrate the big data in the regular production of official statistics, through pilots, exploring the potential of selected big data sources and building concrete applications

ESSNet Big data I

- *Period*: February 2016 - May 2018 (28 months)
- *Number of Partners*: 22 / *Number of WP*: 10
- *Domains*: webscraping job vacancies, webscraping enterprise characteristics, smart meters, mobile phone data, early estimates, multiple domains, methodology

ESSNet Big data II

- *Period*: November 2018 - December 2020 (26 months)
- *Number of Partners*: 28 / Number of WP: 12
- *Domains*: online job vacancies, enterprise characteristics, smart energy, tracking ships, process and architecture, financial transactions data, earth observation, mobile networks data, innovative tourism statistics, methodology and quality, smart statistics

ESSNet Big Data II

WP L – Preparing Smart Statistics

Istat

WP L – Preparing Smart Statistics

- Duration: M12: November 2018 – October 2019

- 11 countries involved :
  AT, BG, FI, FR
  IT, NL, NO
  PL, PT, UK

- Main goals:
  - to explore the IoT in order to produce trusted smart statistics (TSS)
  - to give an overview on the new and relevant topics for official statistics
  - to give recommendations for possible follow-up studies

WP L – Preparing Smart Statistics - Tasks

1. Smart Farming (DE, AT, PL)
2. **Smart Cities** (BG,DE,FR,**IT**,UK)
    1. Case study 1: Varna Limitless: building an IoT system in Varna for bettering the infrastructure and getting statistical data (BG)
    2. **Case study 2: Overview of the H2020 project** (DE,**IT**)
    3. **Case study 3: Combining sensor data and socio-economic data: characteristics of people exposed to air pollution** (FR,**IT**)
3. **Smart Devices** (**IT**, PT, NL)
4. Smart Traffic

Task 2 – Smart Cities

Smart City

A **Smart City** is an urban area that uses different types of electronic Internet of Things (IoT) sensors to collect data and then use insights gained from that data to manage assets, resources and services efficiently. This includes data collected from citizens, devices, and assets that is processed and analyzed to monitor and manage traffic and transportation systems, power plants, utilities, water supply networks, waste management, crime detection, information systems, schools, libraries, hospitals, and other community services (*Wikipedia-EN*)

Smart City

Smart city: city characterized by the integration of knowledge, structures and technologically advanced means of communication and information, aimed at sustainable growth and improving the quality of life (*encyclopedia Treccani*)

A **smart city** is a place where traditional networks and services are made more efficient with the use of digital and telecommunication technologies for the benefit of its inhabitants and business (*European Commision*)

Smart City

Smart City is a set of urban planning strategies aimed at optimizing and innovating public services and at relating the infrastructure of cities to capital human, thanks to the widespread use of new communication technologies, mobility, the environment and energy efficiency, in order to improve the quality of life and meet the needs of citizens, businesses and institutions. (Wikipedia-IT)

Different definitions, but the **common aspects** are:

- benefits for citizens, enterprises and institutions
- the use of ICT to improve the urban services
- intense use of data collected by smart sensors (IoT)

Task 2 - Case study 2

Lighthouse Projects H2020

H2020 Lighthouse projects

- The purpose of the Case Study 2 was to provide a ***first overview*** of the different "smart solutions" that are partly still developed and partly already implemented in cities across Europe
- As a reference, we started from ***12 projects funded by the European Commission*** that aim to bring together cities, industry and citizens to find solutions and business models that can be ***scaled*** and ***replicated*** in other cities
  https://ec.europa.eu/inea/en/horizon-2020/smart-cities-communities
- We carried on a general overview of the 12 projects and identified the main ***domains of interest***:
  - Energy
  - Mobility
  - Infrastructure / ICT

Smart Cities and Communities lighthouse project

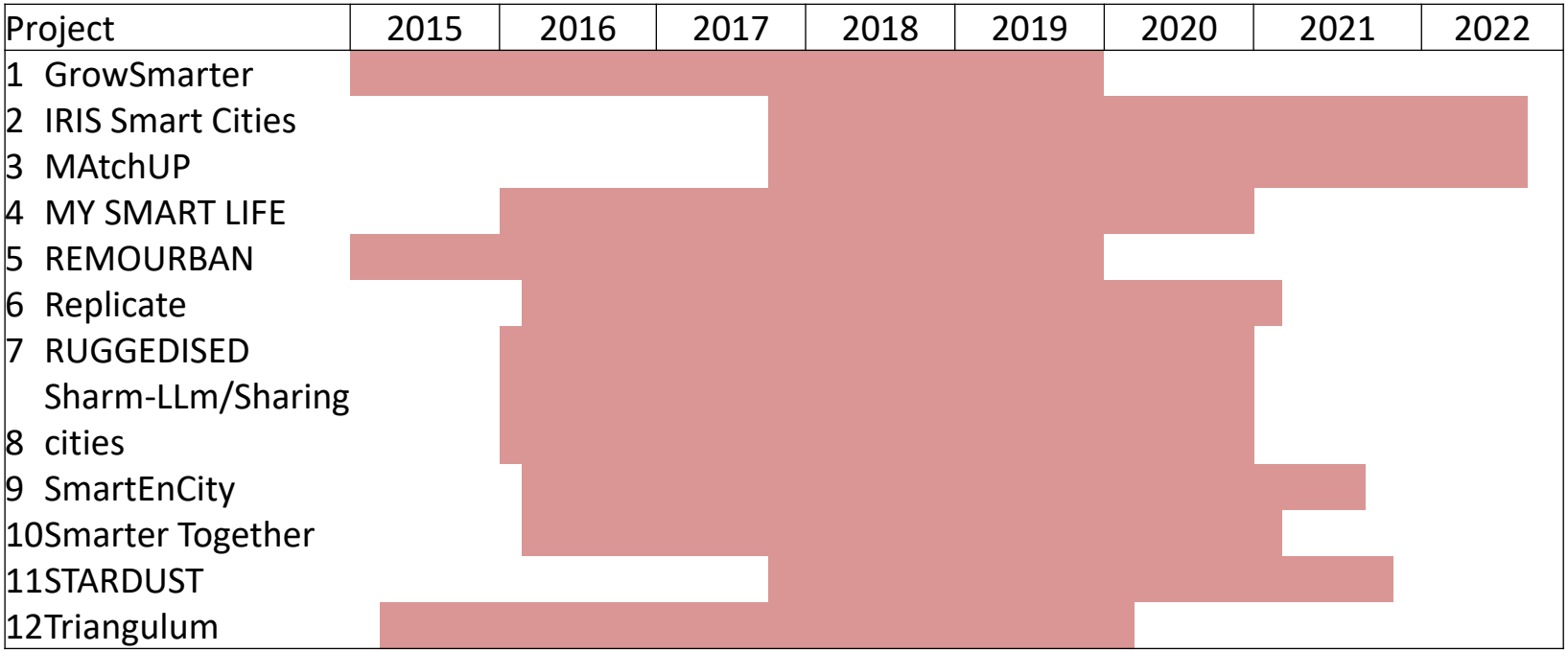
List of the 12 projects:

- Remourban www.remourban.eu
- Triangulum www.triangulum-project.eu
- GrowSmarter www.grow-smarter.eu
- SmartEnCity www.smartencity.eu
- Replicate www.replicate-project.eu
- Smarter Together www.smarter-together.eu
- SharingCities www.sharingcities.eu
- Ruggedised www.ruggedised.eu
- MySmartLife www.mysmartlife.eu
- IRIS www.irissmartcities.eu
- MAtchUp www.matchup-project.eu
- Stardust www.stardustproject.eu

	IRIS Smart Cities https://irissmartcities.eu	MAtchUP http://www.matchup-project.eu	Replicate https://replicate-project.eu/	STARDUST http://stardustproject.eu/
Duration	01.10.2017 - 30.09.2022	01.10.2017 - 30.09.2022	01.02.2016 -31.01.2021	01.10.2017 - 30.09.2021
Contact (Person)	Project coordinator: Haye Folkertsma Gemeente Utrecht (Contact Form)	Project coordinator: Ernesto Faubel Ayuntamiento de València efaubel@valencia.es Communication Secretariat: info@matchup-project.eu Costanza Caffo & Veronica Meneghello Fondazione iCons – iCube programme	Contact Form	Project coordinator: Florencio Manteca CENER – CIEMAT FOUNDATION fmanteca@cener.com Communication: Giulio Mazzolo communication@stardustproject.eu
Objectives	• Energy efficiency • Sustainable mobility • Integrated ICT infrastructures	• Energy efficiency in buildings an renewables • Sustainable mobility and logistic solutions • Integrated ICT infrastructure	• Saving energy consumption • Sustainable mobility • Integrated ICT infrastructures	• Saving energy and Reduce the greenhouse gas emissions by 63%. • Offer greener transportation • Creating and deploying open city information platforms
Light-house cities	Utrecht (Netherland) Nice (France) Gothenburg (Sweden)	Valencia (Spain) Dresden (Germany) Antalya (Turkey)	1. San Sebastián (Spain) 2. Florence (Italy) 3. Bristol (UK)	1. Pamplona (Spain) 2. Tampere (Finland) 3. Trento (Italy)
Follower cities	1. Vaasa (Finland) 2. Alexandroupolis (Greece) 3. Santa Cruz de Tenerife (Spain) 4. Focsani (Romania)	1. Ostend (Belgium) 2. Herzliya (Israel) 3. Skopje (FYROM) 4. Kerava (Finland)	Essen (Germany) Lausanne (Switzerland) Nilüfer (Turkey)	1. Cluj-Napoca (Romania) 2. Derry (UK) 3. Kozani (Greece) 4. Litoměřice (Czech Republic)
Link to project flyer	• IRIS Project Presentation • Info graphics	MAtchUP flyer	Replicate Dissemination Materials	STARDUST Leaflet

Istat

Projects' duration



Kind of data: a double verse

We have analyzed the "Smart City" projects regarding their potential for official statistics, asking us **two main questions**:

- **what kind of data** and what information do **cities need** to turn into smart cities and what data could official statistics, companies, public entities provide them ?
- **what data is generated by smart cities** and how can it be used to produce official statistics?

Data and information

smart data

Data and information the cities need

The cities to become *smart* **need**:

- **statistics and data at the level of small areas** (geo-data encoded at sub-municipal level)
- **statistics and data at family level** (i.e.: administrative registers)

but also **specific information** like for instance:

- information on companies that produce solar and photovoltaic systems that can help them to choose the best technology for developing renewable energy smart system
- data and information useful for identifying areas suitable for exploiting geothermal energy: surface extension, energy potential, number of buildings that can be connected to the geothermal area, etc.

Smart data for official statistics

Energy domain

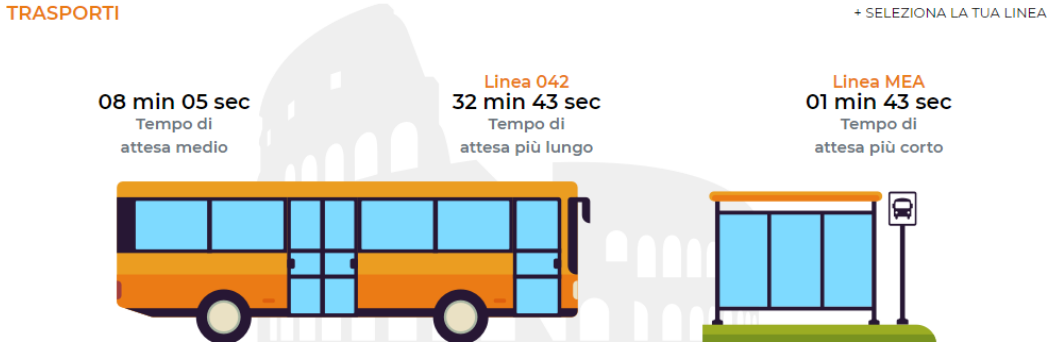
- data of the **smart power management systems**, or **smart metering** (real-time electricity consumption) could be relevant for statistics on energy consumption
- data on electricity consumption generated by **smart devices on E-Vehicles** could be relevant for statistics on emissions

Smart data for official statistics

Mobility domain

- **Smart Mobility Platform** (gathering data from IoT) could be relevant for official statistics, e.g. traffic statistics, mobility statistics, road accidents statistics
- **smart** (LED) **street lighting** with sensors could be relevant for official statistics, e.g. traffic statistics, mobility statistics, road accidents statistics
- **Smart cards** for multimodal transport could be relevant for official statistics on mobility behavior and passenger multimodality travelling

La tua città in tempo reale.



https://www.ilmessaggero.it/smartcity/

Smart data for official statistics

ICT / Infrastructure domain

Data collected and stored in Smart Platforms are surely relevant to produce official statistics

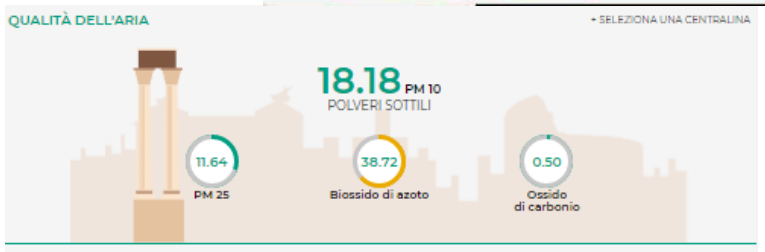
- **monitoring district mobility patterns** could be relevant for official statistics on mobility or traffic statistics
- **real-time data on the amount of solar power or geo-thermal power stored** in the storage systems at districts level could be relevant for official statistics on electricity

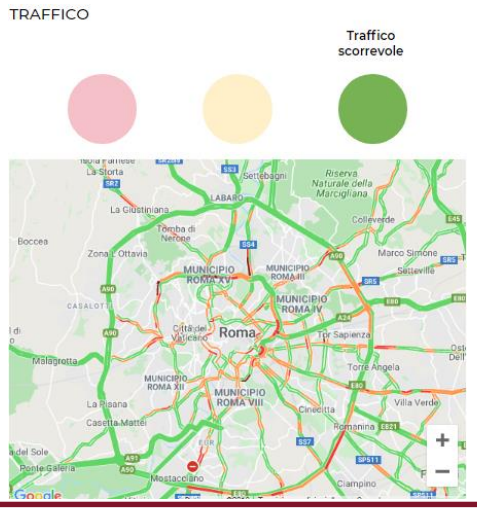
Smart solutions

Smart Data Platforms

They are designed:

- to collect and store data in a standardized format
- to give back information useful to build statistical indicators
- to extract punctual (from a spatial and/or temporal point of view) data through tools and apps
- to provide aggregated information in order to calculate statistical indicators directly





Smart solutions

Smart lamp posts

- can be equipped with many sensors and additional devices to measure pollution, traffic flow, WiFi range extender, etc
- can capture many different kind of data: vehicle traffic, noise and air pollution, water level, pedestrian counting, etc.
- are widhspread on territory with an high capillarity

Case Study

Air pollution

Case study on air pollution (IT,FR)

Inspiration: a study of the European Environment Agency published at the end of 2018 entitled "Unequal exposure and unequal impacts: social vulnerability to air pollution, noise and extreme temperatures in Europe“

This study highlights:

- a global trend towards overexposure of disadvantaged populations
- the lack of information at a fine spatial scale

Case study’s target: to study the socio economic characteristics of people exposed to pollution with the aid of smart sensors and official data about dwellings.

Workflow

1. finding a pilot city equipped with pollution sensors
2. identifying the partner to involve in the project
3. performing an academic literature review, defining the methodology of the study
4. analyzing the data gathered by the sensors
5. analyzing the socio-economic data available at the NSIs and identifying the variables of population and dwellings most useful for the study

Comparison of scoping data

	INSEE	ISTAT
Pilot City	Nice (FR)	Rome (IT)
Population	342 637 inh. (5th in France)	2 873 000 inh. (1st in Italy)
Surface	72 km2	1 285 Km2
Mean household income	15 563 €	37 547 €
Project's stakeholders	• IMREDD (depends from Nice University and the Nice metropolis)	• ARPA Lazio (Regional Agency for Environmental Protection)
Number of pollution sensors	3 fixed + 5 mobile	13 fixed
Pollutant studied	• PM10 • PM2.5 • NO2	• PM10 (Particulate Matter) • NO2 (Nitrogen Biosside) • O3 (Ozone) • CO (Carbon Oxide) • SO2 (Sulfur Oxides)
Temporal frequency	1 hour	1 hour
Temporal period	5 year	1 year
NSI's source	• census data • fiscal data	• census data
Geographical scale available	35 m2	1 Km2

Istat

Data available

Census data on population and dwellings

- 116 variables grouped in: family, immigration, education, employment, population and housing
- geographic level : census section

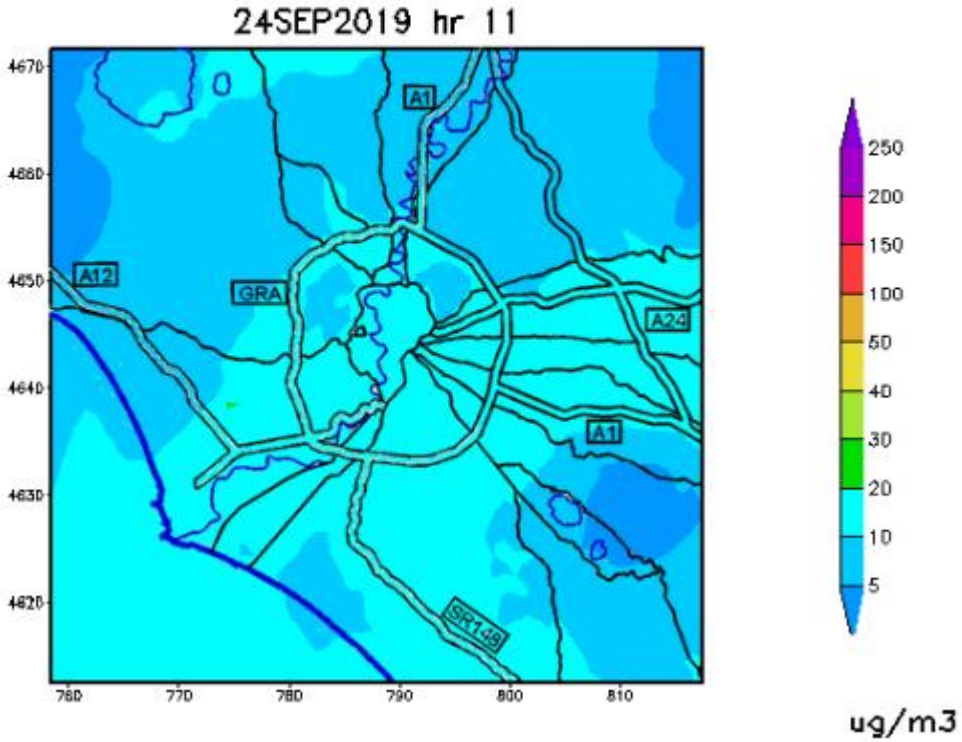
Pollution data

- are referred at main air pollutants: sulfur dioxide (SO2), carbon monoxide (CO), fine particulate matter (PM10 and PM2.5), ozone (O3) and Benzene
- are available at daily, monthly and yearly level
- are collected by 13 official station
- are available “near-time” data in continuous time at a scale of 1 km^2

ARPA Lazio – near real time data

Concentrazione di Particolato 10 micron(PM10)

Simulazione modellistica senza dati assimilati dalle stazioni.



Methodology

- the census data are available at different aggregation levels: the most detailed level is the **census section**.
- for urban areas, the air pollution data are mapped in a **grid of 1 km square**
- the size of the census section **is comparable** with the grid on which the air pollution data are mapped
- in the city of Rome there are about 13,000 census sections and the average area is less than 1 km^2
- the census and air pollution **data** are both **geo-localized**
- with the **GIS** (Geographic information system) **methodology**, we can link the two data sources reporting them at the census section level

In this way, we can elaborate the **average data** of the pollution and the social profiles of the population referred to a **single territorial dimension**

Task 3 – Smart Devices

What is a smart device ?

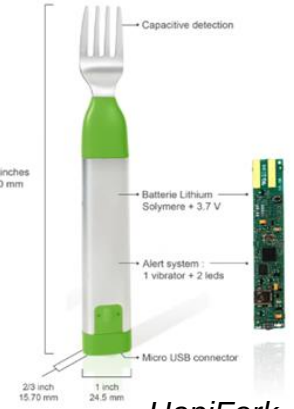
«A **smart device** is an electronic device, generally connected to other devices or networks via different wireless protocols such as Bluetooth, Zigbee, NFC, Wi-Fi, LiFi, 3G, etc., that can operate to some extent *interactively* and *autonomously*.” (*Wikipedia*)

Interactive means that users can influence the behavior of the device

Autonomous means that the device can operate for a longer time without human intervention to do its task

Smart Halo

SmartPlate



HapiFork

Main targets of Task Smart Devices

- to give an **overview** on smart devices and citizen science data from the viewpoint of official statistics
- to explore some pratical use cases

Smart devices long list (2019)

- **Smartphones**
- **Smart home devices**: speakers, TVs, thermostats, kitchen, smoke/AQ, garden, security, pets
- **Smart devices for healthcare**: diagnosis, glucose, blood pressure, heart rate, oxygen level, pulse, alcohol, asthma, inhalers, bioscarf, smart pills
- **Smart devices for fitness**: jogging, cycling, heart rate, oxygen, sleep, brain, diets
- **Smart devices for transport**: traffic lights, cameras, counters, vehicles, bike sharing
- **Smart devices for travel**: suitcase, tracking, transport cards, hotspots
- **Other smart devices**: waste, IO-link

Considerations for definition of use cases

- **types of data access** : direct data access (DA), pushing computation out (PCO), manufacturers portals (MP) and citizen science (CS)

- **classification of the nature of the data source** used to collect data by smart devices: public sector data, private sector data or community data

- **maturity of devices** in terms of number of users and intensity of use

A focus on data access' types

Direct access

- direct from the device; it isn't the most common case, but it's possible that the NSI receives data from the smart devices directly. Example: a smart traffic lights system send their data not only to the traffic regulation authorities, but also directly to the statistical office; this could help create real-time traffic statistics

Pushing computation out (PCO)

- indirect (aggregated) data from device manufacturer. *Example: a mobile network operators run aggregation algorithms designed together with the statisticians and send its aggregated (anonymized) results to the statistical office for further processing*

A focus on data access' types

Manufacturers' portals:

- indirect from portals where device manufacturers publish (usually aggregated) data on the use of the smart device they manage. Example: the portals of companies that produce sports activities devices

Citizen science:

- indirect from an open data citizen science portal. As is frequently in citizens science's projects, citizens tend to do measurements themselves using smart sensors that they connect to one or more citizen science portals that publish the data, usually as open data . A well-known example is the Luftdaten network of air quality measurement. (https://luftdaten.info/en/home-en/)

Smart devices' task: use cases

Use Case 1: Citizen science smart devices

- *Telraam* project: measurement of traffic intensity via a low cost smart cameras or via smartphones; it could represent a starting point for counting other kinds of objects (https://www.telraam.net/en)

Use Case 2: Passenger mobility via the smartphone

- development of a smartphone app to deduct people's mobility pattern using machine learning techniques

Use Case 3: Travel patterns based on smart travel cards

- aggregated data: travel patterns, commuter movements, stages of a trip etc.

Final considerations

A wide variety of topics concerning the subject of IoT have been examinated and the final considerations are:

- follow-up studies are needed
- the data landscape is still in a development state
- **data access** and **availability** are often an issue
- the O.S. must secure the **data protection** (trends: *data anonymization* and *data aggregation*)
- the **confidentiality** should be protected using new technologies and procedures (such as *cryptography* and *blockchains* techniques)
- the use of data generated by IoT tools offers many possibilities to enrich
- the **quality of data** gathered by smart systems and their **representativity** should be examinated carefully
- the use of the ‘new data’ and the adopting a new statistical production process of the NSIs to involve this kind of data, have to be explored in further studies

References

- «Towards a Reference Architecture for Trusted Smart Statistics" - ESS Task Force on Big Data and Official Statistics (Fabio Ricciato, Michail Skaliotis, Albrecht Wirthmann, Kostas Giannakouris, Fernando Reis)

- https://www.eea.europa.eu/publications/unequal-exposure-and-unequal-impacts

- https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/images/5/50/WPL_Deliverable_L2_Description_of_the_findings_regarding_Task_2_Smart_Cities_2019_11_01.pdf

- https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/images/9/91/WPL_Deliverable_L3_Description_of_the_findings_regarding_Task_3_Smart_Devices_2019_10_30.pdf

Thank you for your attention !