SDMA - A Framework for Sensor Data Management and Analysis

Application to Participatory Sensing

david
données et algorithmes
pour une ville intelligente et durable

Karine Zeitouni, Nov 2016

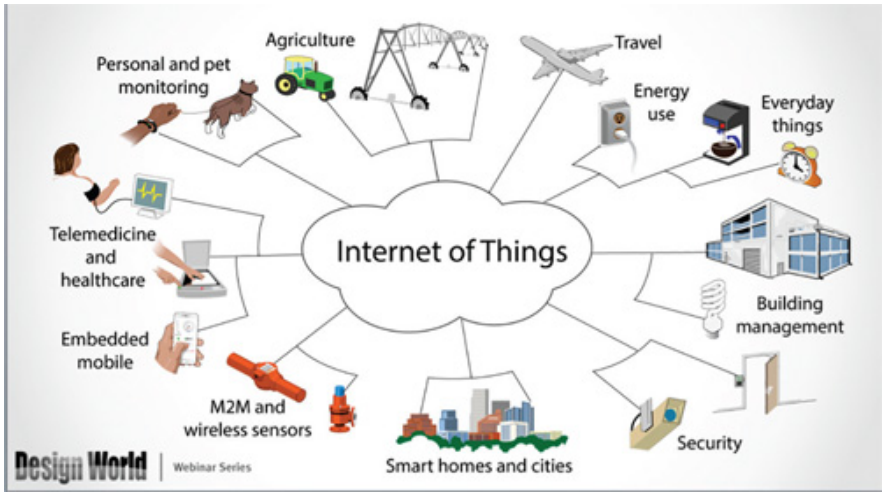
UNIVERSITÉ DE VERSAILLES ST-QUENTIN-EN-YVELINES
université PARIS-SACLAY

Context – Sensor Data Analysis

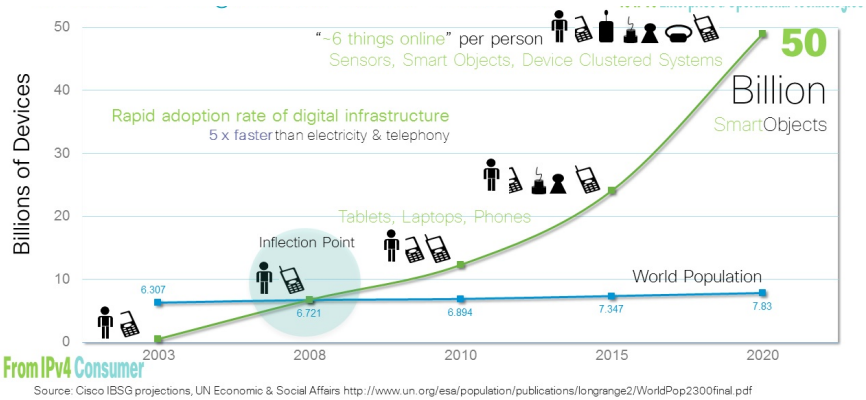
- IoT is "*a pervasive and ubiquitous network which enables monitoring and control of the physical environment by collecting, processing, and analyzing the data generated by sensors or smart objects*"

- Specificity of the data
  - *Series of <time[, measures]>*
  - *Multivariate time series + metadata*
  - *Big & noisy data (incomplete & uncertain)*

- Analysis requirements:
  - Study and implement multivariate time series (MVTS) analysis in the context of IoT,
  - Apply on real or realistic datasets and evaluate the performance and the scalability
  - Leverage (and contribute in) open source libraries, like scikit-learn



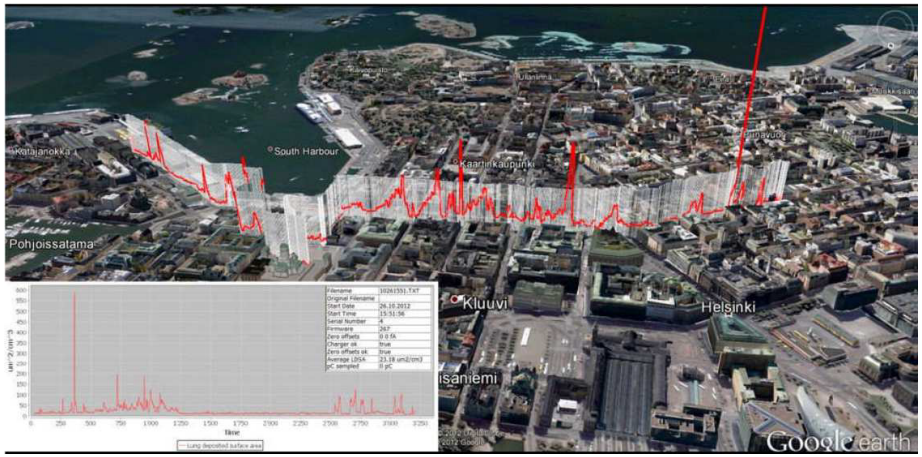
Variety of IoT applications

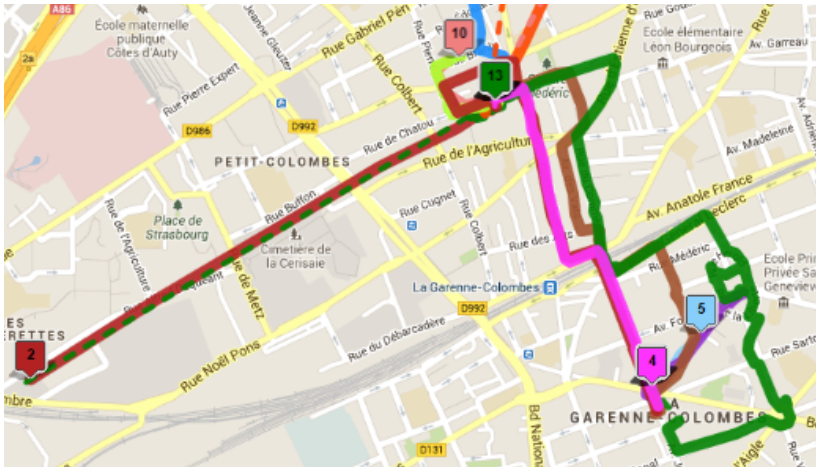


Growth of the IoT (http://www.cisco.com)

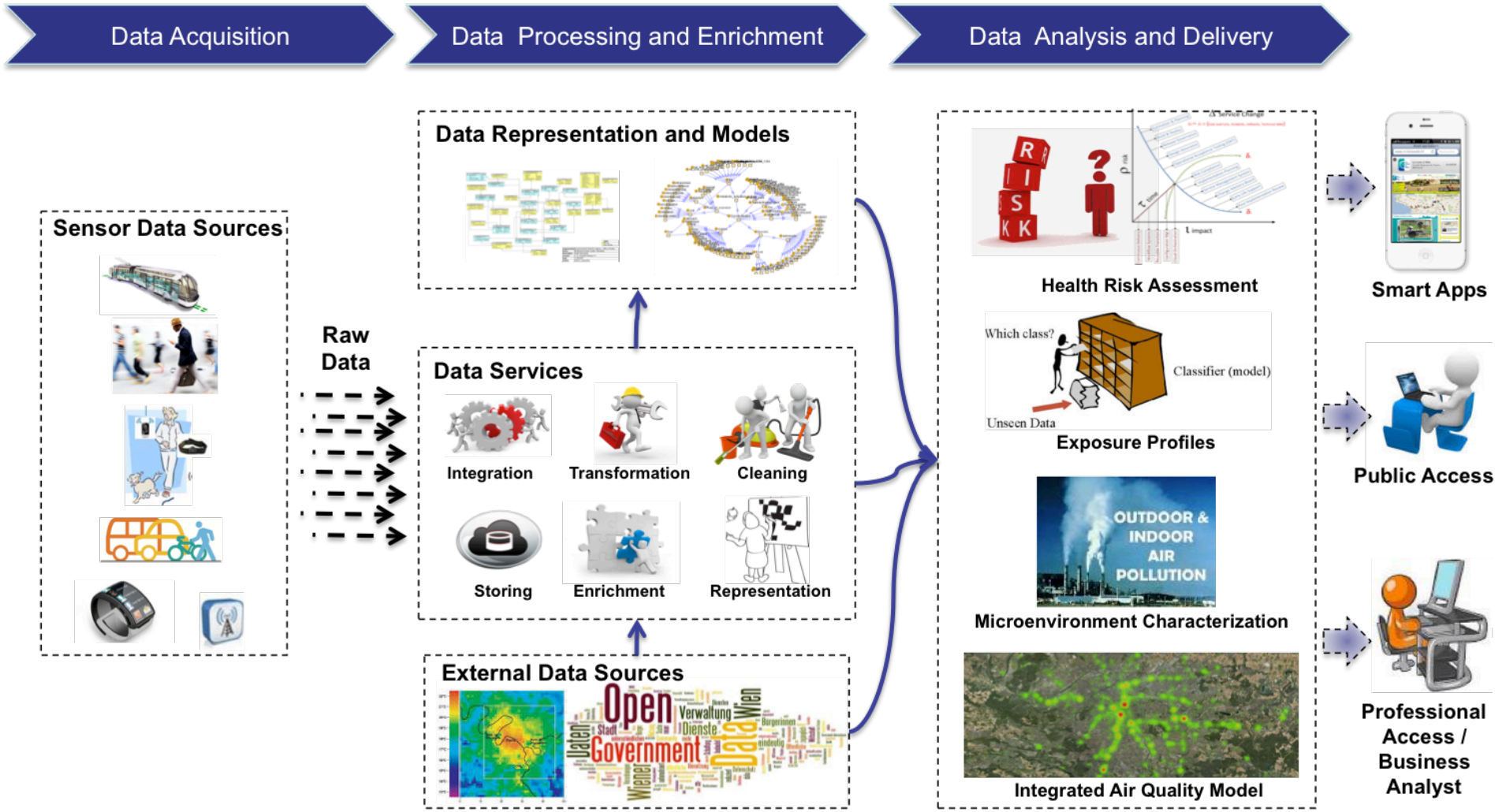
Use Case
Community-based exposure sensing

- Polluscope (ANR project) will collect, operate, and analyse data of real individual's exposure to air pollutants and the impact on his/her health.
  1. Based on the personal lightweight connected environmental and health sensors
  2. Covers both indoor and outdoor daily life activities, such as transport, work, dwellings, …
  3. May complement the fixed-sensors based air quality assessment





Polluscope Data Workflow



Data Collection Campaign (2017-2019)

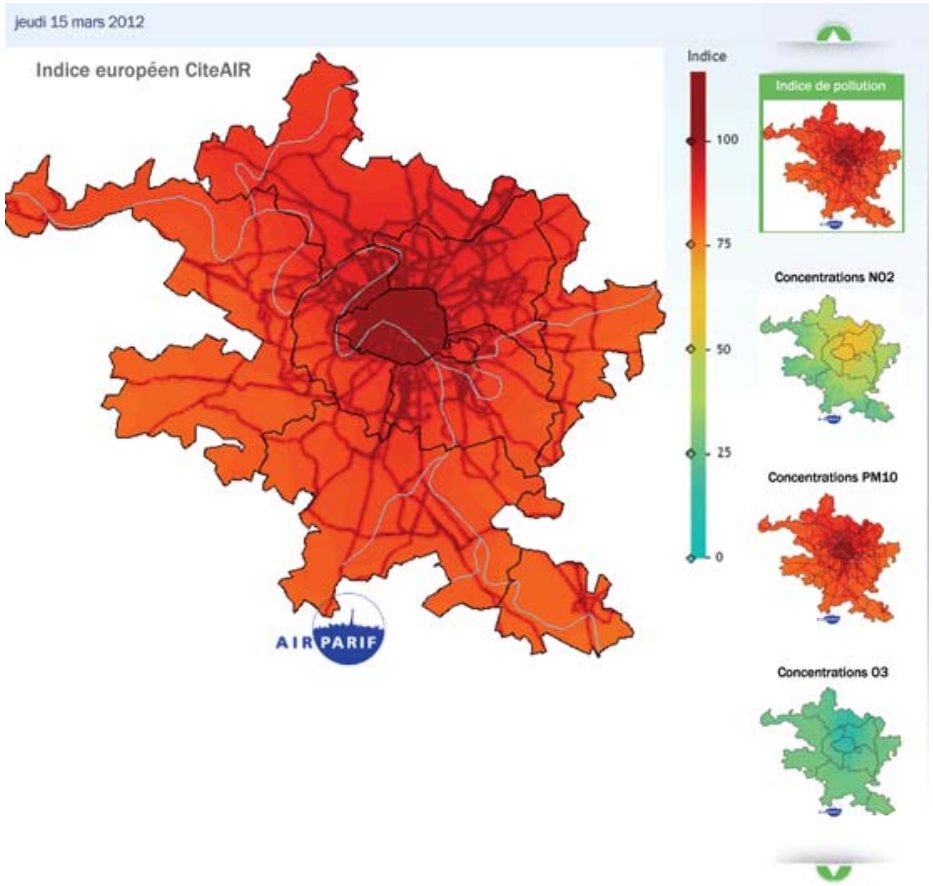
- Participants will be equipped with a multi-sensor box containing:
  - a GPS receiver, an accelerometer, humidity, temperature, Ozone, NO2, O3, PM10, PM2.5, black carbon , Volatile Organic Compounds (VOC)
  - + respiratory and cardiac sensors

- 150-200 participants will be recruited in 2 years – viewed twice a year (Winter & Summer times):
  - General population study to validate the participatory sensing scheme
  - An epidemiological study, involving patients suffering from asthma or COPD and healthy participants (from RECORD Cohort) matched with the hospital patients according to age, sex, socioeconomic status, etc.

Since the campaign starts in october 2017, there is a need of « simulation »

Proposal - Input data sources(*)

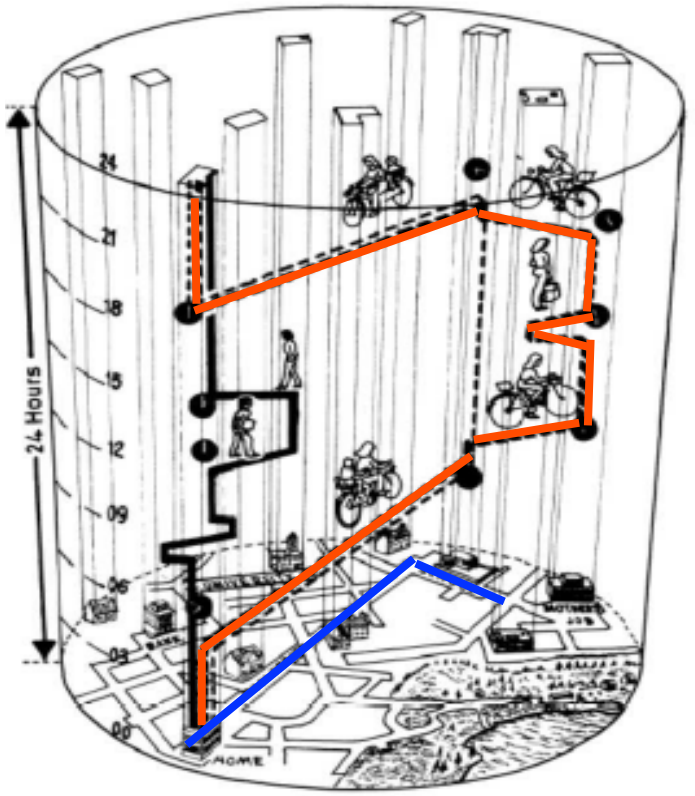
Air quality data (Airparif)



Exemple de cartes un jour de pollution soutenue
Ces cartes journalières sont disponibles sur l'ensemble de l'agglomération pour la veille

Activity/Mobility data (*Record Study - INSERM*)

record-study.org



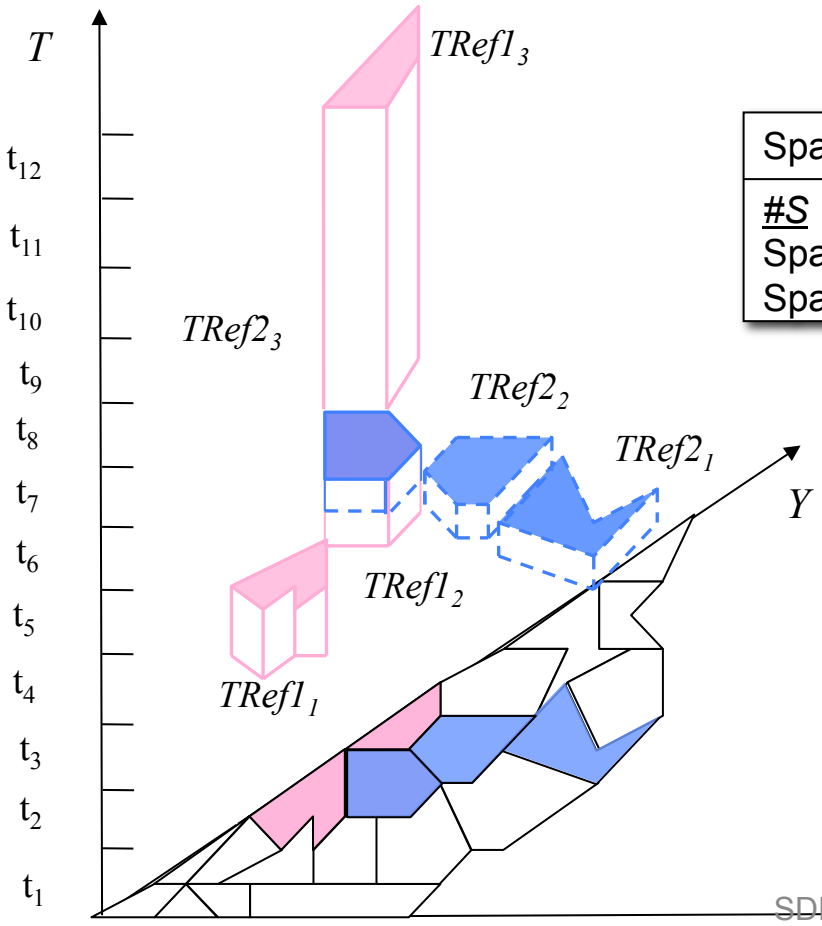
(*) Subject to an agreement

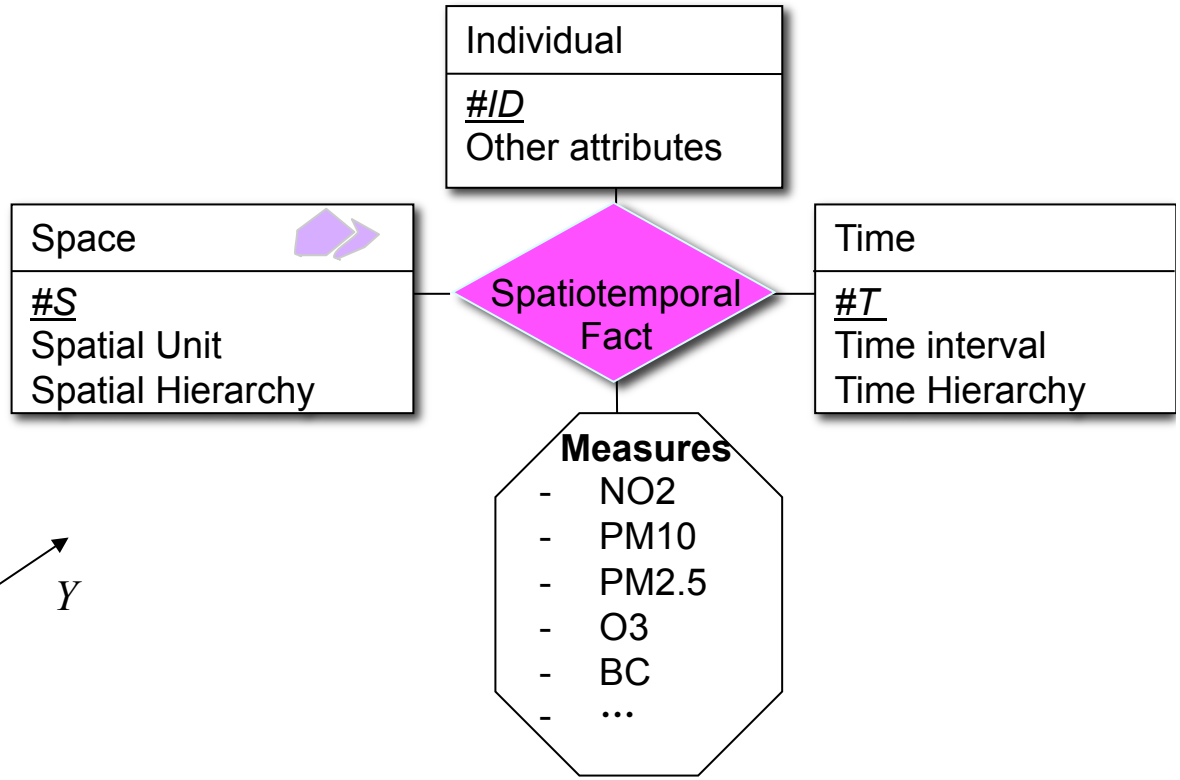
1st task - Data preparation

- Use of real mobility data from Record study
  - RECORD GPS : Feb 2012 – June 2013 (7500 trips)
  - RECORD with health sensors : Sept 2013 – June 2015 (9000 trips)
- Match them according to GPS locations + time with real air quality maps of AIRPARIF
  - Estimate the indoor air quality (e.g., NO2) by using transfert functions
  - Anomymize personal data by applying well-known location anonymisation techniques

✓ This will estimate / simulate the crowd sensed data, targeted in Polluscope

2nd Task - Multi-dimensional data model





3rd Task - Data Analysis

- Examples of multi-dimensional analysis & data mining:
  - Characterize the exposure per microenvironment (near the traffic, indoor, per mode of transport, …)
  - Characterize the exposure per period of time
  - Characterize the exposure per category of users (age, gender, activity type, etc.)
  - Identify profiles of exposure / (clustering, Functional Data Analysis ?)
  - Relate individual's exposure to major air pollutants and her cardiorespiratory health status
  - Relate individual's exposure to her mobility / activity profile
  - Comparison / enrichment of the existing air quality modelling approaches and moving sensor data heat-maps.
  - Predict short-term exposure based on the real-time data and the predicted activity / mobility ….

✧ *Adressing issues of scalability and real-time processing for data stream*

Social and Economic Impacts

- IoT data analysis is valuable for:
  - In the case of manufacturing: reducing maintenance costs, avoiding equipment failures and improving business operations.
  - In the environmental health case:
    - better knowledge of individual risks of pollution, decision support, and behavioural change.
    - The participants can access personalized exposure data, which provides a good incentive and may impact their bahaviour
- Economic benefits:
  - Pollution costs 15% to 30% of the annual health system deficit in France (1 and 2 billion per year) only for cardiorespiratory morbidity due to pollution.
  - The users, especially patients suffering from COPD or asthma, could be alerted proactively when a risk is predicted, and allow them to reduce their exposure.

More infos about Polluscope

http://www.david.uvsq.fr/polluscope

Multi-disciplinary Partnership