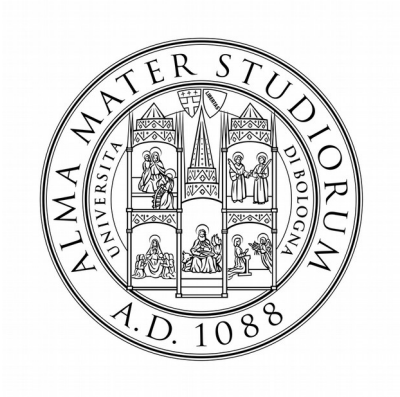


Federico Montori, Luca Bedogni, Luciano Bononi
Department of Computer Science and Engineering (DISI)
University of Bologna, Italy

{federico.montori2, luca.bedogni4, luciano.bononi}@unibo.it

On the Integration of Heterogeneous Data Sources for the Collaborative Internet of Things

Presented by: Federico Montori

IEEE 2nd Internal Forum on Research and Technologies for Society and Industry, Technologies for smarter societies (RTSI)

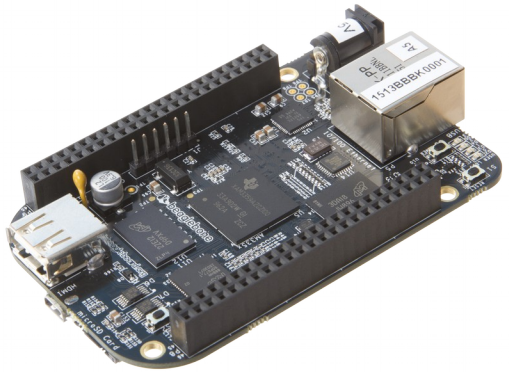
Bologna, 7,8,9 September 2016

Track: IoT and Big Data: Efficient Management of Sensor/Cloud Integration.

IoT and DIY crowdsense

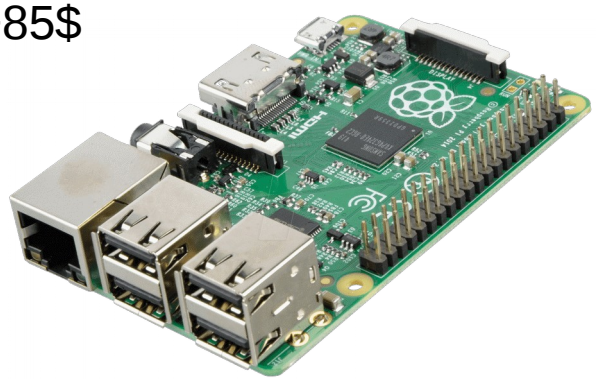
IoT faced an enormous growth during the last years.

Makers can build their own sensing solutions at a lower and lower cost and can upload easily the produced data to an Open Data platform such as ThingSpeak.

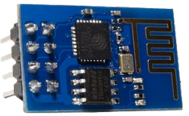


BeagleBone Black ~85$

Arduino UNO ~28$



Raspberry PI ~50$



ESP 8266 ~10$

Open Data growth

Open data are a possible solution for privates and companies willing to share their useful sensed data.

Data is released either by privates (non-reliable) or environmental agencies and institutions (reliable) through data platforms in which sensed data is updated periodically.

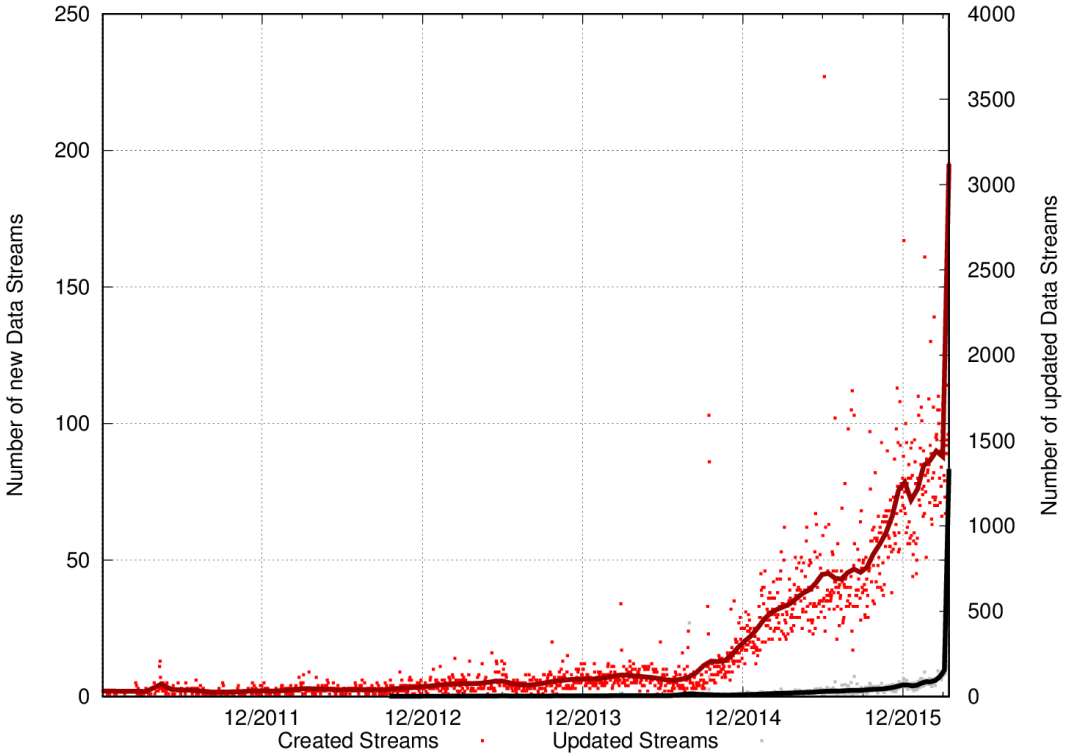
arpa
agenzia regionale prevenzione e ambiente dell'emilia-romagna

OpenSignal

ThingSpeak

sparkfun ELECTRONICS

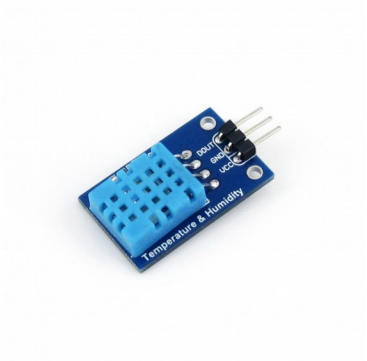
As an example: ThingSpeak data streams



ESP 8266 first version

ESP 8266 programmable through Arduino SDK

Collaborative IoT



DHT 11
Temperature and
Humidity Sensor

~ 4 $

What if users want to retrieve sensed data in order to build their own services? Sometimes they cannot afford home made and/or ad-hoc solutions:

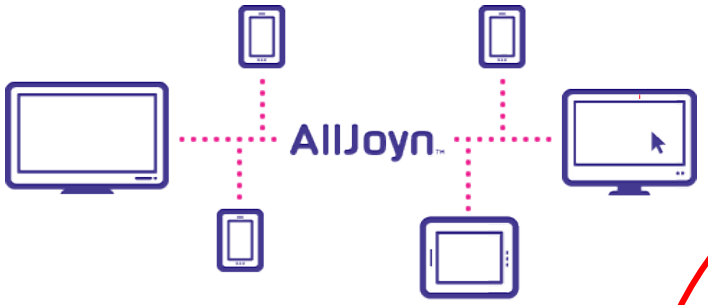
- High cost of some sensors
- Inability in dealing with complex development
- No infrastructure

Kanomax PS-2
Pollen Sensor

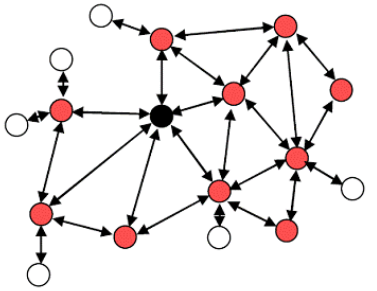
~ 1980 $

We propose the Collaborative IoT to cope with such restrictions.

Isolated ecosystems



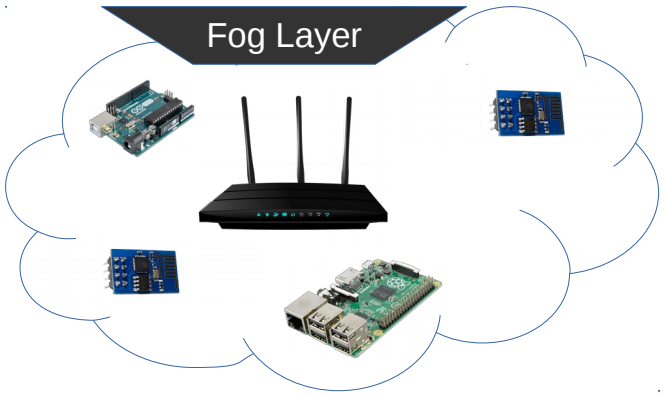
AllJoyn Network



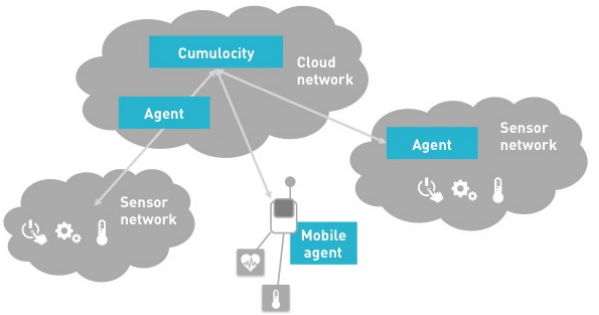
Ad-Hoc ZigBee Network

Existing ecosystems are hardly able to communicate between each other due to protocol heterogeneity and different data structure.

We propose an unification platform using a standard raw data structure.



DIY Home Automation WiFi Network



Cumulocity Network

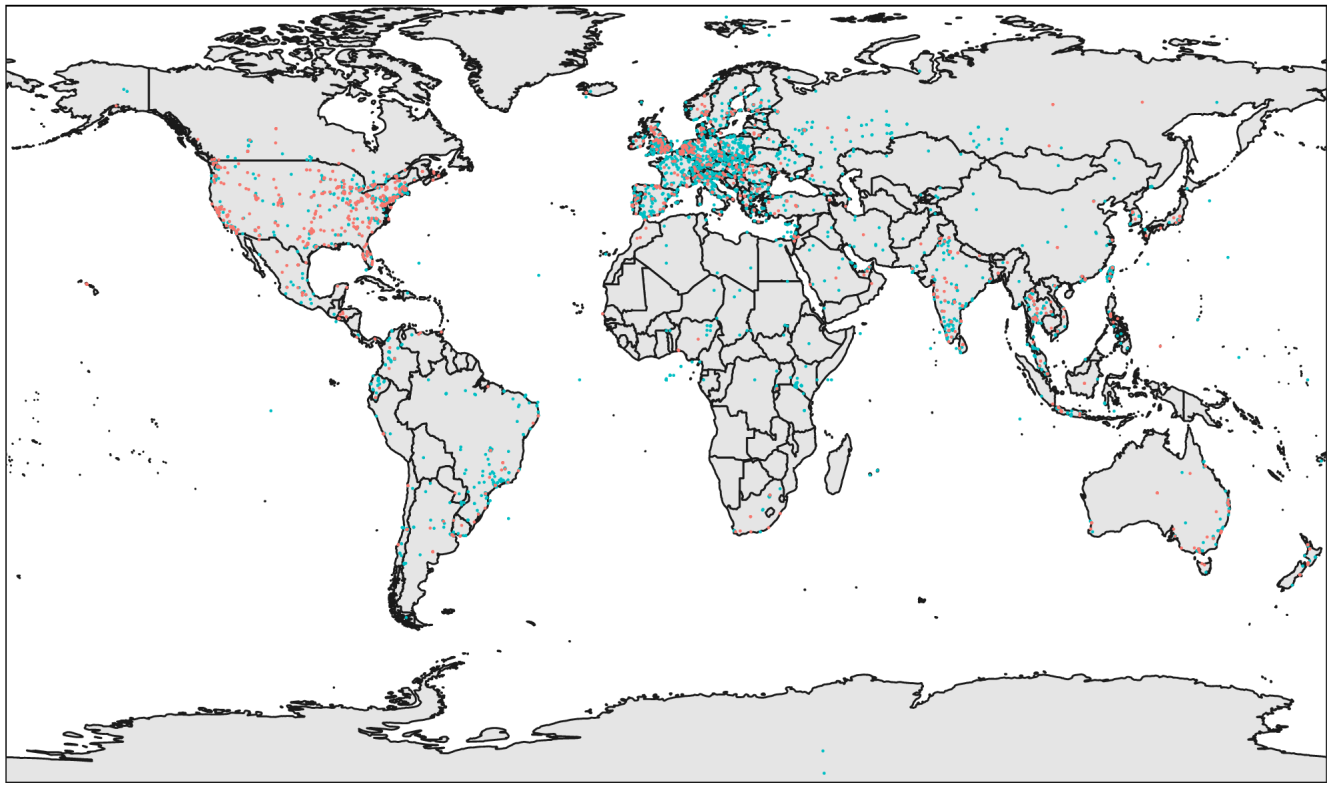
Our Aim

In order to face the amount of connected devices forecasted, we need to provide an easy accessibility to raw data and services for the end users.

Within the scope of standardization of Collaborative IoT using open data we propose:

- Cooperative paradigm.
- Standardized raw data structure.
- Example of data unification coming from open data platforms.
- An adaptable architecture, including any role for end users.

Meet ThingSpeak and SparkFun



ThingSpeak:

- Open source data platform
- 28806 public streams
- ~15% is geolocalized
- More popular in Europe

SparkFun:

- Microcontroller seller
- 3575 public streams
- ~62% is geolocalized
- More popular in United States

Importance of merging sources:

- Increase coverage
- Increase sampling number
- Increase fields of application

Data Unification

- **Stream ID:**
  in TS is a progressive number, in SF a string of 20 random ASCII characters.
- **Stream Name:**
  in both platforms is decided by the owner and might carry useful information about the stream.
- **Geolocalization:**
  in TS is given in GPS coordinates, in SF in GPS coordinates of the center of the city.
- **Tags:**
  user-defined tags facilitating the categorization of data.
- **Creation Timestamp:**
  when the stream was created; in some cases the first updates have been erased.

1..N

- **Field Name:**
  User-defined name for the measurement.
- **Value:**
  Value of the measurement
- **Timestamp**

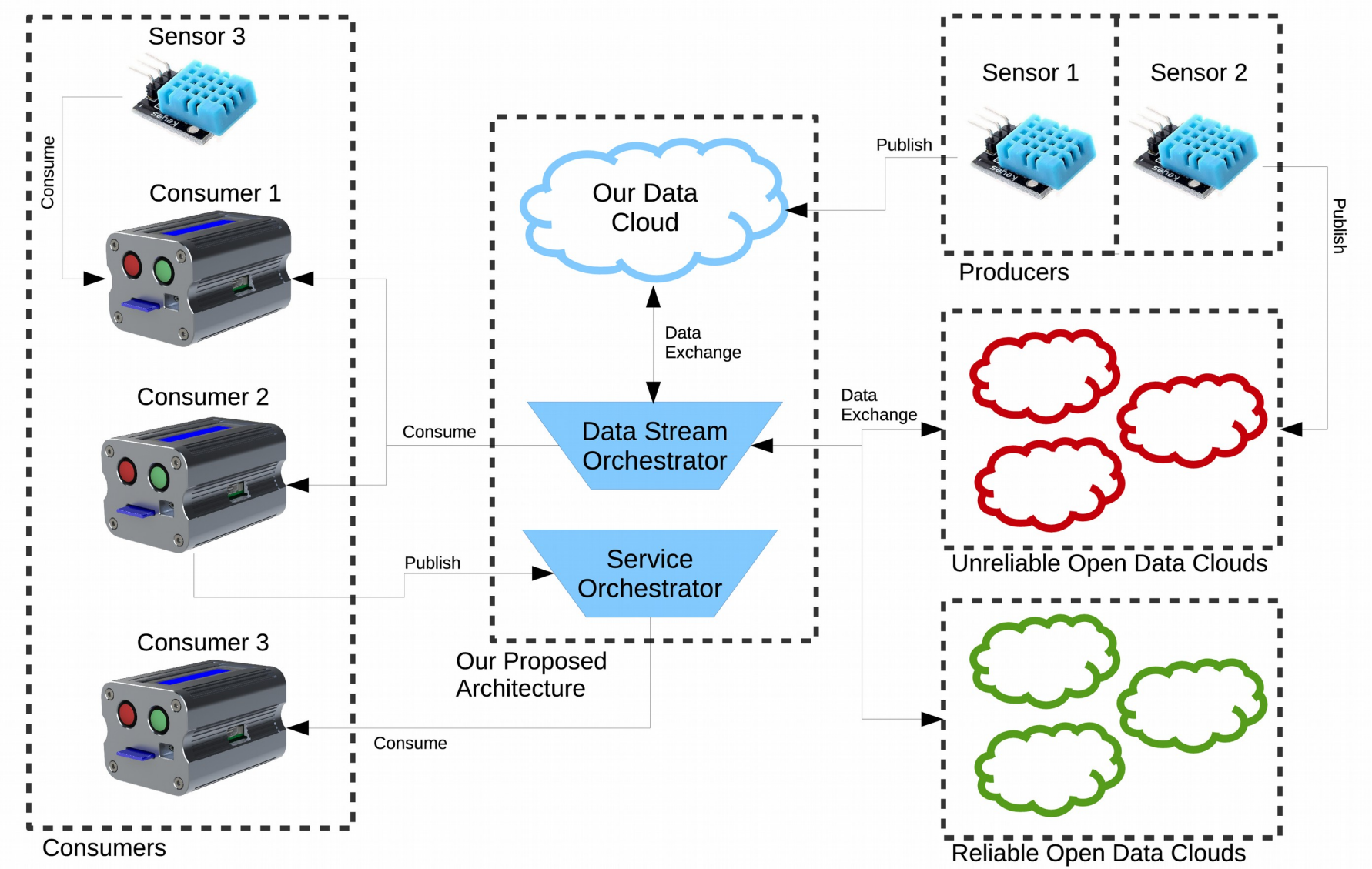
- **Class?**

We created a unified database with the common information from the two platforms.

- **Precision?**

- **Area of Validity?**

Architectural Model



User publishing on Open Data Platforms **/** User publishing in our Data Platfom **/** User aggregating data as a service

User consuming from his private sensor **/** User consuming from raw data orchestrator **/** User consuming a service

Simple Example application

Owned Temperature Sensor

Non-owned near Pollen sensor

It's hot

High concentration

Low concentration

Open the window not facing the sun

Turn on air conditioning

Both electricity consumption and allergy safety are considered.

Conclusions

In this paper we have studied the challenging topic of data integration between heterogeneous data sources for the Internet of Things.

We analyzed the differences, and proposed a new architecture to integrate them together, along with the ability to deliver custom made services to the end users.

Future Work

- Machine learning algorithm to classify sensed data in order to facilitate data search.
- Integrate governmental and more specialized data.
- Implement efficient orchestration capabilities.
- Incentives.