

Giacomo Chiesa, Dipartimento DAD, Politecnico di Torino

giacomo.chiesa@polito.it

Abstract. L'interesse verso tecnologie e strategie per la costruzione delle smart city e degli smart building favorisce la diffusione di soluzioni ad alto contenuto ICT che spesso utilizzano grandi quantità di dati. Trattando il tema del monitoraggio urbano, non si può precludere dal guardare alle innovazioni introdotte dai bigdata, soprattutto per quel che riguarda il tema della "datizzazione", passando dalla raccolta di un numero limitato di campioni all'accumulo di quanti più dati possibile, a prescindere dai futuri utilizzi. Il paper si focalizza sulla fase di produzione di dati dal monitoraggio di variabili ambientali tramite stazioni di misura diffuse sul territorio, ai fini di identificare problematiche e possibili soluzioni operative per la costruzione bottom up di bigdata.

Parole chiave: Big data, Monitoraggi ambientali, Smartcity, Modello e Progetto, Datizzazione

«La vera rivoluzione non sta nelle macchine che elaborano i dati, ma solo nei dati in sé e nel modo in cui li usiamo». (Mayer-Schönberger & Cukier, 2013)

Introduzione

I dati e le informazioni sono oggetto di un cambiamento in merito alle modalità di organizzazione e sviluppo della scienza che, come ricorda Nielsen (2012), permettono di affermare che «stiamo reinventando la scoperta». Inoltre, la creazione di una SmartCity, strategia per un futuro urbano su cui si concentrano i crescenti interessi dell'Unione Europea e dell'Italia, prevede, tra le altre azioni, un forte interessamento nei confronti delle tecniche di digitalizzazione della città, intese come spazi di ibridazione tra il mondo reale e quello virtuale. La città intelligente, che per sua natura dovrebbe risultare coesa, connessa e innovativa, necessita di informazione e di modelli, i quali permettono quella misura dell'ordine e del disordine, o meglio dell'entropia urbana (Wiener, 1966), nonché una progettazione intelligente. La creazione, l'utilizzo e la diffusione dei dati e delle informazioni deve

Data, BigData and smart cities. Considerations and case study on environmental monitoring

Abstract. The growing interest in technologies and strategies for constructing smart cities and smart buildings promotes the spread of ICT solutions which often use large amounts of data. Nowadays, urban monitoring are often interrelated with the innovations introduced by BigData and the neologism "datization", passing from the collection of a limited number of datapoints to the accumulation of as much data as possible, regardless of their future uses. The paper focuses on the production phase of data from the monitoring of environmental variables by using several measurement stations spread on the territory. The aim is to identify operational problems and possible solutions for a bottom-up construction of BigData datasets.

Keywords: Big data, Environmental monitoring, Smartcity, Model and Design, Datization

relazionarsi con i processi urbani e architettonici complessi in un flusso bidirezionale, divenendo parametri di progetto e civilizzazione. In questo contesto i modelli e gli strumenti di gestione e progetto sono interessati da una costante innovazione grazie al progressivo passaggio da un regime di scarsità di dati ad un'abbondanza di informazioni generate da reti di sensori e attuatori connessi in real time. Le tecniche di creazione, elaborazione e utilizzo dei dataset sono, quindi, direttamente interfacciate con la gestione della complessità delle reti e delle città intelligenti. Tuttavia, il passaggio al mondo dei bigdata comporta nuove implicazioni e necessita di azioni di re-engineering e re-design.

In quest'ottica si è sviluppato un progetto di ricerca volto a verificare problematiche e criticità nell'utilizzo di stazioni di monitoraggio low cost e DIY per la costruzione di database di variabili ambientali alla scala urbana. Le fasi della ricerca sono la costruzione di un corpo teorico-interpretativo delle implicazioni delle nuove tecniche di datizzazione e costruzione di scenari urbani basati su dati diffusi e real time; l'identificazione delle problematiche connesse alla costruzione di modelli di analisi basati su dati spesso carenti per quel che riguarda le metainformazioni; la definizione di aspetti legati alla produzione e all'utilizzo di dataset e datapoint; lo sviluppo di alcuni nodi sperimentali di una possibile rete di monitoraggio di natura bottom up.

Il contesto

Il rapporto uomo-tecnologia sta attraversando una fase di cambiamento a seguito delle innovazioni e delle nuove implicazioni delle ICT (tecnologie di informazione e comunicazione) nella vita quotidiana (es. Floridi, 2013; Braham and Hale, 2007; Barker and Erickson, 2005; Ocelli and Staricco, 2002).

«The real revolution is not in the machines that process data, but only in the data itself and the way in which we use them» (Mayer-Schönberger & Cukier 2013)

Introduction

Data and information are changing following new organization and development of sciences, which, as noted by Nielsen (2012), allow us to say that «we are reinventing discovery». In addition, the creation of a SmartCity, a strategy for cityfutures that reaches a growing interest from the European Union, provides, among other actions, a strong interest in the techniques of digitalization of cities, which become spaces of hybridization between the real world and the virtual one. The smart city, which by its nature should be cohesive, connected and innovative, requires information

and models, which allow the measure of order and disorder –urban entropy (Wiener, 1966)–, and a smart design and planning. The creation, use and dissemination of data and information relate to urban and architectural complex processes in a two-way flow, and become parameters of design and civilization. In this context, models and design tools have to be constantly innovated in order to face the gradual shifting from information scarcity to information abundance generated by networks of sensors and actuators connected in real-time. Creating, processing and use of datasets are action directly connected with the management of the complexity for networks and smart cities. However, the bigdata transition requires innovative actions and instruments of re-engineering and re-design.

The paper presents the results of a

Questo cambiamento radicale ha dirette implicazioni in ambito architettonico, come già sottolineato nel pluricitato articolo di David Celento, "Innovate or perish", che ricorda come the «architects' refusal to embrace technological innovations invites their extinction» (Celento, 2007).

Le ripercussioni indotte dalle ICT sui processi e sulle attività umane possono essere organizzate utilizzando quattro macro assi, così come riportati da Floridi (2013):

1. la riduzione dei confini tra mondo reale e mondo virtuale (Sakamoto et al., 2008; Sass & Oxman, 2006; Oxman, 2006; Mitchell, 2005; Milgram & Colquhoun, 1999; Negroponte, 1995);
2. l'ibridazione tra mondo naturale e mondo artificiale (Hochberg et al., 2012; Chiesa, 2010; Bar Cohen, 2006; Hashimoto and Dijkstra, 2004; Benyus, 1997);
3. il passaggio dalla scarsità all'abbondanza di informazioni (Mayer-Schönberger and Cukier, 2013; Nielsen, 2012; Xu, 2012; Weinberger, 2012; Shirky 2010; Wiener, 1966);
4. il passaggio dal primato dell'entità al primato dell'interazione (City Form Lab; Weinberger, 2012; Barabasi, 2003).

I quattro assi si articolano in azioni di virtualizzazione/datizzazione del reale e di materializzazione del virtuale. Questi due processi opposti trovano spazio di applicazione nel concetto di piattaforma che può essere intesa come spazio di gestione, alle diverse scale, dei progetti complessi e della smartness e che può configurarsi come un nodo o una rete poiché rappresenta sia lo spazio del processo sia la creazione di modelli. Secondo la definizione di Jeff Rothenberg (1989) «modeling is one of the fundamental processes of the human mind» (Rothenberg, 1989) e, inoltre, «a model represents reality for the given purpose; the model is an abstraction of reality in the sense that it cannot re-

research project designed to assess problems and critical issues in the use of low-cost and DIY (Do It Yourself) monitoring stations for the construction of databases of environmental variables at the urban scale. The research focuses on the following phases: the definition of a body of theoretical and interpretative implications of new techniques for datization and construction of urban scenarios based on real-time flow of information and bigdata; the identification of issues related to the construction of analysis models based on data that didn't present enough meta-information; the analysis of main aspects related to the production and use of datasets and datapoints; the development of experimental nodes for testing a monitoring network for a bottom-up bigdata production.

Background

The relationship between man and technology is decisively changing, primarily because of the rapid development of ICTs (Information and Communication Technologies) and their implications in everyday life (e.g. Floridi, 2013; Braham and Hale, 2007; Barker and Erickson, 2005; Occeilli and Staricco, 2002). This radical change has direct implications in architecture, as already reported by David Celento in an article entitled "Innovate or perish", in where He remembers that the «architects' refusal to embrace technological innovations invites their extinction» (Celento, 2007).

It is possible to classify the effects induced by ICTs on human activities using four main axes, as reported by Floridi (2013):

1. the reduction of boundaries be-

present all aspects of reality». Ogni modello è caratterizzato da tre attributi essenziali "reference", "purpose" e "cost-effective" (Rothenberg, 1989).

Tuttavia, se il modello (reale o virtuale) è strumento di progetto e intermediario tra le idee degli architetti e i costruttori (Sass and Oxman, 2006) almeno sin dal rinascimento (Maldonado, 1992), sono le tecnologie modellistiche attuali che permettono la gestione della complessità progettuale contemporanea. Da un lato, infatti, nuove modalità di produzione e gestione del legame tra progetto e materia permetteranno di gestire gradi elevati di complessità e di rispondenza modello-reale (Celento, 2007), dall'altro il progredire della rivoluzione dei big data modificherà le modalità di costruzione dei dataset in un'ottica in cui le informazioni in sé, oltre alle tecnologie, saranno oggetto di innovazione e sviluppo.

I modelli possono essere utilizzati nelle smart city per ottimizzare alcune funzioni urbane a partire dalla raccolta di dati derivanti dal mondo reale e raccolti in real time. L'uso di dati real time in grande quantità permette, almeno in parte, di spostare la virtualizzazione e l'uso dei modelli scientifici alla fase di elaborazione, arricchendo la qualità e la velocità dei processi e garantendo una migliore rispondenza con il mondo reale basata su un'ottica di operational rating piuttosto che di asset rating.

Problematiche applicative dei bigdata

La principale problematica legata alla proliferazione di nodi di misurazione diffusi consiste nella mancanza di metadati legati alle misurazioni, ovvero quelle informazioni che introducono ad esempio la tipologia di variabile trattata (tipo di dato), la fonte, la definizione operativa delle

tween the real world and the virtual one (Sakamoto et al., 2008; Sass & Oxman, 2006; Oxman, 2006; Mitchell, 2005; Milgram & Colquhoun, 1999; Negroponte, 1995);

2. the hybridization between the natural world and the artificial one (Hochberg et al., 2012; Chiesa, 2010; Bar Cohen, 2006; Hashimoto and Dijkstra, 2004; Benyus, 1997);

3. the shift from scarcity to abundance of information (Mayer-Schönberger & Cukier, 2013; Nielsen, 2012; Xu, 2012; Weinberger, 2012; Shirky 2010; Wiener, 1966);

4. the transition from the primacy of nodes to the primacy of interactions (City Form Lab; Weinberger, 2012; Barabasi, 2003).

The four axes are based on the virtualization/datization of reality and the materialization of virtuality. These two opposite processes could find a

place for application in the platform, concept that can be defined as a space to manage, at different scales, complex projects and urban smartness. It can be seen as a node or a network because it represents the space where processes are managed using models. According to the definition of Jeff Rothenberg (1989) «Modeling is one of the fundamental processes of the human mind» (Rothenberg, 1989) and, moreover, «a model represents reality for the given purpose; the model is an abstraction of reality in the sense that it cannot represent all aspects of reality». Each model is characterized by three essential attributes "reference", "purpose" and "cost-effective" (Rothenberg, 1989).

Models (real or virtual) are used as design tools and interfaces between the ideas of architects and builders (Sass & Oxman, 2006) at least since

variabili e quant'altro. Il metadato è un elemento fondamentale per passare dal dato all'informazione. Inoltre, considerando la natura georeferita o in ogni caso posizionale delle misurazioni, è importante considerare la dipendenza areale delle misure, per quanto sia soprattutto fondamentale per i piccoli numeri quando si procede con la normalizzazione dei dati. Nel caso dei big data, la grande ampiezza del dataset permette di limitare le problematiche di natura areale trattando le scelte di aggregazione dei dati a diversi livelli territoriali per verificare l'insorgere di criticità. Infine, come ricordano Mayer-Schönberger e Cukier (2013), nelle analisi basate sui big data, i singoli datapoint non sono più elementi critici poiché il regime di abbondanza di informazioni riduce l'importanza del dato singolo rispetto all'analisi venendo meno la sua capacità di inquinare il risultato finale.

Tuttavia, riguardo la qualità dei dati si deve tenere in considerazione:

- la completezza (copertura dell'entità coinvolta rispetto all'area di studio e mantenimento delle informazioni necessarie in fase di classificazione);
- la consistenza logica (preservazione delle relazioni logiche tra i dati, evitando contraddizioni tra contenuti informativi e garantendo coerenza e affidabilità);
- l'accuratezza (precisione metrica del dato);
- l'accuratezza posizionale, per dati georeferiti;
- l'accuratezza temporale (invecchiamento del dato e precisione nell'identificazione del momento di presa);
- l'accuratezza tematica (discrepanza tra tema monitorato e tema reale);
- la risoluzione;
- il lineage (tracciabilità temporale delle modifiche successive).

the Renaissance (Maldonado, 1992). However, the innovative modeling technologies enable a new management of complexity in contemporary design. On the one hand, it is possible to handle high degrees of complexity and compliance model-reality (Celento, 2007) by using new methods of production and management between design and materialization; on the other hand the bigdata revolution changes the technologies of construction of datasets suggesting that the information itself is a subject of innovation and development.

Models can be used in smart city to optimize some urban functions by collecting high amount of data in real time and from the real world. The use of real-time data in large quantities could allow moving the virtualization process and the use of scientific models to the stage of processing, increas-

ing the process quality and ensuring a better correspondence with the real world based on operational rating rather than asset rating databases.

Application problems of bigdata

A big problem related to the proliferation of monitoring tools and nodes consists in the lack of metadata of measurements, in other words that information that describe, for example, the type of measured variable, its source or its operational definition. Metadata are a key element to move from data to information.

Furthermore, it is important to consider the areal dependence of measures, related to data geopositioning, which is a crucial aspect especially for normalized small numbers. For bigdata, the large size of datasets limits this problems and it is possible to aggregate data at different territorial

Un'ulteriore spinosità è legata all'ontologia del dato, la quale è utile «a classificare e a esplicitare i caratteri di ciò che classifica» (Ferraris, 2003). È fondamentale riflettere sulle modalità di rappresentazione della conoscenza in funzione delle scelte di archiviazione, gestione, interoperabilità, trasferibilità, accessibilità dei dati (es. Rivoltella, 2010; Odifreddi, 1994). L'ontologia racchiude una schematizzazione concettuale (rappresentazione di un modello di un dominio dato) di natura esplicita e non ambigua, formalmente espressa in un linguaggio conosciuto secondo una conoscenza condivisa. In una modalità di elaborazione dati basata su database strutturati, l'ontologia dei dati e delle variabili è fondamentale per poter procedere con l'analisi e il confronto tra le informazioni. È anche fondamentale che i dati siano omogenei e ordinati per poter essere utilizzati e processati. Tuttavia, gli strumenti che operano su grandi dataset, come ad esempio Hadoop, non hanno necessariamente bisogno dell'operazione di estrazione trasferimento e carico dei dati per l'analisi. In altre parole «presume che i dati non siano omogenei e ordinati – anzi, si assume che siano troppo ingenti per poterli ripulire prima di processarli» (Mayer-Schönberger and Cukier, 2013).

Alle tematiche sin qui accennate, occorre aggiungere una serie di problematiche di natura software e hardware, tra le quali si ricorda soprattutto l'interoperabilità dei modelli utilizzati all'interno delle differenti piattaforme di gestione ed elaborazione. La possibilità di accedere ai dati non riguarda soltanto la trasferibilità degli stessi da e verso piattaforme differenti, ma anche il persistere nel tempo delle modalità di accesso e di utilizzo (aggiornamenti, release successive, cambi di interfaccia e strumenti).

Nella progettazione di una piattaforma di monitoraggio diffuso si ritiene essenziale garantire nel tempo:

levels to check the occurrence of criticality. Finally, as reported by Mayer-Schönberger & Cukier (2013), in a bigdata analysis, individual datapoints are not critical elements because the abundance of information reduces the ability of polluting the final result by single data.

However, it is important to take into account the following issues for assuring the quality of data:

- completeness (coverage capacity of dataset in comparison to the entire population and preservation of information during the process of classification);
- logical consistency (preservation of logical relationships between data, avoiding contradictions between information content and ensuring consistency and reliability);
- metric accuracy;
- geopositional accuracy;

- temporal accuracy (data obsolescence and correct identification of intake time);
- thematic accuracy (discrepancy between monitored and real data);
- resolution;
- lineage (temporal tracking of changes).

Another criticality is represented by the ontology of data, which is useful «to classify and explain the characteristics of the content of classification» (Ferraris, 2003). It is essential to correctly choose the types of representation of knowledge according to the settled methods of storage, management, interoperability, portability and accessibility of data (e.g. Handgun, 2010; Odifreddi, 1994). The ontology contains a conceptual, explicit and unambiguous scheme (representation of a model of a given domain) that is formally expressed in a known

TAB. 1 | Parametri richiesti per la compilazione della scheda. Lo sviluppo di un database implementabile dovrebbe agevolare la compilazione fornendo un elenco di componenti già riconosciuti
List of parameters required for register a metro station. The development of a database should be implemented to facilitate the compilation providing a list of pre-defined components

- la connettività tra gli hub e la piattaforma anche nel caso di aggiornamenti;
- le modalità di interoperabilità e funzionamento del flusso di dati dalla produzione all'utilizzo/visualizzazione;
- il funzionamento e la tipologia di interfaccia, per evitare di richiedere alla singola utenza costosi e complicati aggiornamenti;
- l'utilizzo, anche al variare della formattazione, degli scenari e dei dataset in un'ottica di estensibilità dei data point raccolti.

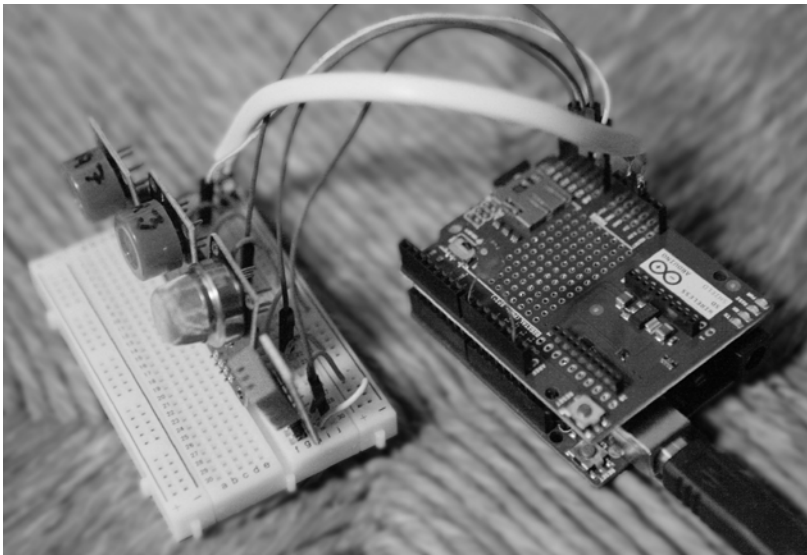
Monitoraggi urbani nelle città intelligenti

Sulla base delle considerazioni effettuate si suggerisce una possibile modalità organizzativa e un relativo esempio applicativo di una piattaforma di monitoraggio urbana volta anche a utenti bottom up. Tale modalità propone la compilazione di schede, proposte come strumento atto a incrementare la conoscenza sui singoli dati prodotti, contribuendo alla costruzione di metadati e all'utilizzo delle informazioni. Lo studio si concentra su tecniche per la produzione di dati specifiche per monitoraggi diretti (es. di variabili ambientali) e, in que-

sto, si differenzia da altre analisi legate ai big data basate sulla dataizzazione di fonti indirette (quali ad esempio Twitter, Facebook, Google). Lo scopo consiste nel costruire una base comune che ogni utente può direttamente implementare nel sistema tramite maschere di inserimento. Si costituisce, quindi, una community di stazioni per il monitoraggio, ognuna delle quali permette di conoscere un minimo di ontologia dei dati trasmessi e comporta una definizione dei metadati connessi alle variabili lì monitorate. I singoli nodi possono, ovviamente, essere costruiti per il solo utilizzo nel monitoraggio o possono integrarsi con altre funzioni proprie del nodo locale quali sistemi domotici e altre componenti elettroniche domestiche o pubbliche sparse sul territorio, che potrebbero essere rilocalizzati nel tempo. Le piattaforme elettroniche descritte permettono di monitorare una grande quantità di variabili modificando i sensori impiegati e i conseguenti schemi software e hardware. Il fine ultimo delle stazioni è la produzione diffusa di dati in modo tale che ogni nodo venga descritto dall'utente sulla base dell'elenco riportato in (Tab. 1).

FIELDS FOR DESCRIBING A MONITORING NODE
Starting date and hour of measurement
(possible Switching off date and hour)
Type of timing measurement used (RTC modules, PC data, other solutions)
Geopositioning of monitoring station (on a map, using coordinate values or by GPS shield). For mobile station (e.g. on a vehicle) has to be equipped with a GPS module
Building type
Type of microcontroller board (with specifications)
Transmission modality/ies
Storage modality/ies
Other shields or function (e.g. actuators)
No. of sensors
Upload - hardware scheme (e.g. by using Fritzting)
Upload – photos of the node and localization
Upload – software definition
FIELDS RELATED TO THE CONNECTED SENSORS
Measured variable/s
Type and manufacturer (ID e.g. DHT11 – with specifications)
Power supply
Connection Pin/s
Delay between measurements
Type of datapoint (single or averaged value – on a fixed number of values or moving average)
Transmission and store technology
Unit of measure and formula (from resistance to this unit)
Placement (indoor or outdoor)

TAB. 1 |



01 | Uno dei tre nodi di monitoraggio al momento dell'assemblaggio hardware, foto di G. Chiesa
One of the three monitoring nodes during the assembly, photo by G. Chiesa

TAB. 2 | Descrizione dei nodi di monitoraggio utilizzati per il caso studio
Description of monitoring nodes used for the case study

Simulazione di un monitoraggio. Descrizione, analisi e risultati

contesti indoor destinati a diverse destinazioni d'uso: un edificio residenziale, un edificio adibito ad uffici e un edificio destinato a laboratorio universitario (Tab. 2). Le stazioni sono state sviluppate per verificare il funzionamento di sensori di qualità dell'aria e presentano caratteristiche analoghe per poter confrontare i dati registrati in contemporanea dai diversi nodi (Fig. 1). Sono stati impiegati tre sensori della serie MQ e un sensore di T e UR (DHT11) per ogni stazione. Per facilitare l'importazione in Excel, si è costruita a livello software una variabile nominale capace di riportare le misure dei differenti sensori, inoltre si è

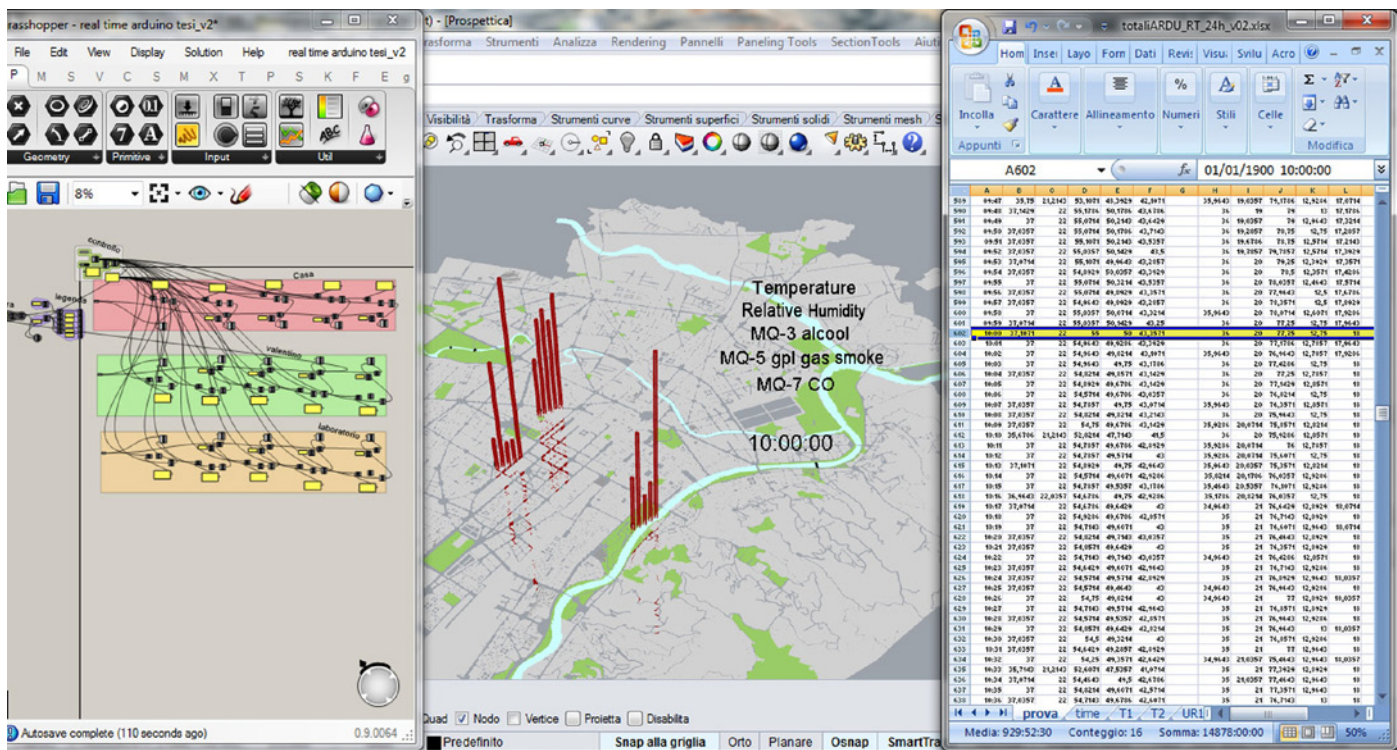
Nel corso della ricerca è stata condotta una simulazione di monitoraggio basata su tre stazioni di presa dati localizzate in diverse posizioni in Torino in

proceduto ad effettuare una media dei valori registrati nel corso di un minuto per quanto questa operazione potrebbe essere efficacemente processata adattando il software scritto per la scheda Arduino. L'obiettivo della simulazione è dimostrare come sia possibile sviluppare sistemi di monitoraggio diffusi DIY capaci di popolare una mappa real time grazie ai dati registrati dalle tre stazioni (5 variabili a stazione). Le visualizzazioni sono effettuate ricorrendo all'impiego del software Rhinoceros 5 e dei plug-in Grasshopper e Ghowl usando come base lo shapefile fornito dal Comune (SIT) (Fig. 2).

Lo script compilato in Grasshopper permette di aumentare il numero delle stazioni, ma richiede un'elevata potenza di calcolo del PC. Inoltre, l'aggiornamento dei dati può essere gestito usando un software di connessione automatica porta seriale-foglio di calcolo come ad esempio PLX-DAQ della Parallax, Inc.

TAB. 2 |

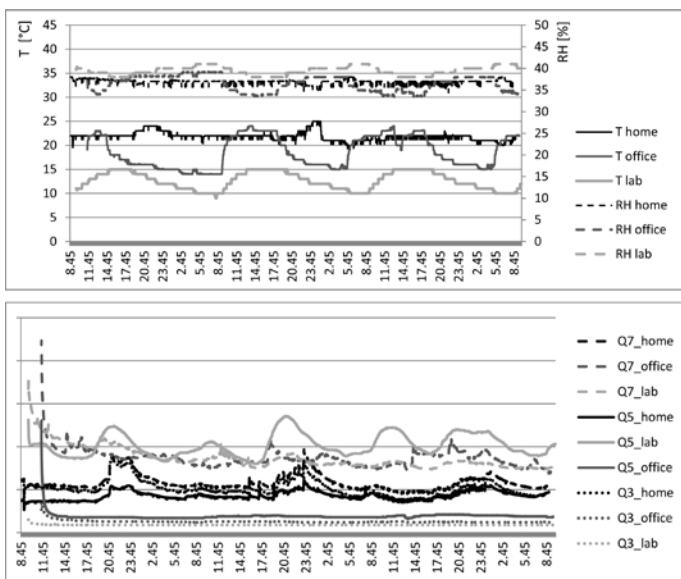
Monitoring Node	I_	II_	III_	Ethernet
Geopositioning	45°04'51"N 7°39'17"E	45°03'12"N 7°41'06"E	45°03'44"N 7°39'24"E	45°03'12"N 7°41'06"E
Location	Indoor – Residential	Indoor – Office	Indoor – Laboratory	Indoor – Office
Microcontroller Board	Arduino Uno R3	Arduino Uno R3	Arduino Uno R3	Arduino Uno R3
Sensors	Mq 3; Mq 5; Mq 7; Dht I I	Mq 3; Mq 5; Mq 7; Dht I I	Mq 3; Mq 5; Mq 7; Dht I I	Figaro Tgs2602, Figaro Tgs2442 Dfrobot Lm35
Power Supply	Usb	Usb	Usb	Usb
Data Storage & Transmission	Pc – Serial Port/Usb	Micro Sd – Stand Alone	Micro Sd – Stand Alone	Ethernet Shield + Micro Sd
Data & Clock Ref.	Pc Clock Using Plx-Daq	Microcontroller Clock Ref.The Starting Point	Microcontroller Clock Ref.The Starting Point	
Starting Point	08:45 Day I	11:33 Day I	09:46 Day I	
Ending Point	09:06 Day Iv	09:40 Day Iv	10:25 Day Vii	



02 | Rappresentazione geolocalizzata dei dati monitorati dalle tre stazioni alle ore 10:00 del secondo giorno di misurazione
 Georeferred representation of the monitored data from the three stations at 10:00, second day of measurement

L'analisi dei dati permette di leggere diversi fenomeni in base al cambiare dei valori monitorati, quali ad esempio l'apertura di una finestra, l'ingresso in casa degli abitanti, lo spegnimento del riscaldamento nell'ufficio, i fumi della caldaia dell'edificio di fronte, l'andamento periodico del traffico veicolare (Q5) nella zona del laboratorio tracciato dagli inquinanti rilevati (Fig. 3). La temperatura dell'abitazione mostra i picchi serali dovuti alla presenza degli abitanti e alle attività di cottura dei cibi, al con-

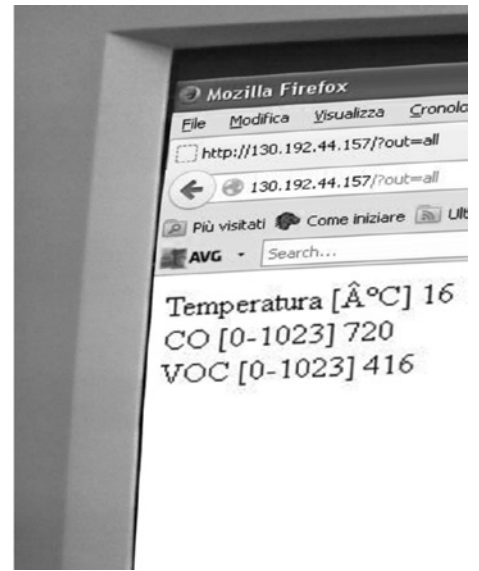
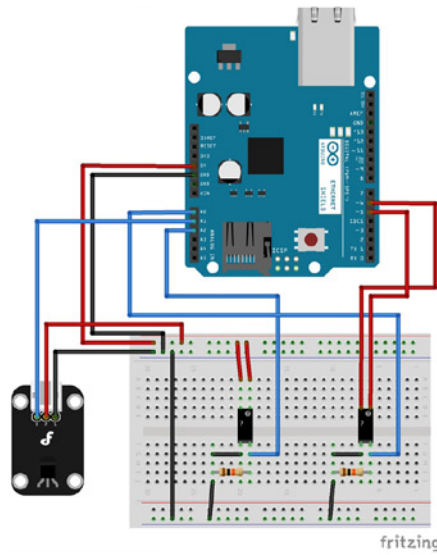
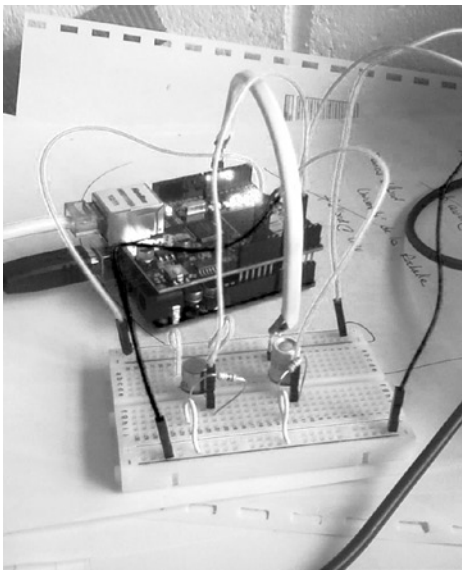
trario l'andamento del riscaldamento in laboratorio è di natura periodica sulla base di un ciclo temporale prefissato. Infine la temperatura dell'ufficio è soggetta ad una gestione dei ventilconvettori che risente di un ciclo di accensione-spegnimento dell'impianto e della regolazione manuale da parte degli occupanti. Si è successivamente testato il funzionamento di un nodo differente capace di trasmettere via ethernet i risultati di monitoraggio in tempo reale.



03 | Andamento dei valori rilevati dalle tre stazioni
 Graph of measured data from the three stations

language according to a shared knowledge. Ontology of data and variables is essential for analyzing and comparing information in a structured dataset. Order and homogeneity of data are other essential issues to use and process information collected in database. However, several tools that operate on large datasets, such as Hadoop, do not necessarily need to extract, transfer and order data before analyze them. In other words Hadoop «assumed that data are not homogeneous and ordered - indeed, it considers that datasets are too large to be possible for a program to clean them up before use» (Mayer-Schönberger and Cukier,

2013). Furthermore, it is necessary to consider a series of problems concerning software and hardware and, in particular, the interoperability of models between different management platforms and analysis processes. This topic relates not only to the transferability of datasets to and from different platforms, but also to the possibility over time to access and use them (updates, new releases, changes in interface and tools). In order to design a good platform for analyzing and storing monitored data from several different stations, it is essential to ensure over time:
 - the connectivity between hubs and



04 | Il nodo predisposto per il collegamento in ethernet e visualizzazione dei valori monitorati su un terminale di rete
 Photo and breadboard view of the Ethernet node. On the right, a monitor showing the monitored values in real time using LAN

L'obiettivo specifico di questa seconda esperienza è validare una diversa modalità di trasmissione dati.

Il nodo è stato organizzato intorno a tre sensori differenti rispetto all'esempio precedente: un sensore di temperatura e due sensori di gas (CO e TVOC) (Fig. 4).

La lettura dei risultati o delle informazioni trasmesse tramite Ethernet può essere effettuata da qualsiasi dispositivo alla rete. Collegando l'uscita USB/seriale del nodo con un computer è possibile leggere, grazie alla programmazione software effettuata, il numero di utenti che sta visualizzando quanto trasmesso dal nodo, il funzionamento della scheda SD di stoccaggio e le informazioni inerenti la connessione (MAC, IP, gateway e subnet). Lo sviluppo della stazione di monitoraggio ha reso possibile verificare il funzionamento di modalità di trasferimento dati in real time tramite reti locali o internet.

platform even in case of upgrades;

- the method of operation and interoperability of data flow from production to use / display;
- the operation and the type of interface, in order to avoid expensive and complicated upgrades of each nodes, especially in bottom up platform;
- the use of datasets and scenarios allowing the extensibility of collected datapoints.

Smart city and urban monitoring

Previous considerations are used for proposing a platform for urban monitoring, and a case study, able to connect single users in a bottom up and open scenario. This platform is based on a monitoring-node data sheet, which becomes a tool for describing each collected data and for collecting metadata and operational information. The research focuses on

the techniques for data producing in direct monitoring (e.g. environmental variables), and, for this reason, it differs from other analyses related to big data, which are based on datization from indirect sources (such as Twitter, Facebook, Google).

The objective is to build a common ontology in order to allow each user to directly implement the system using input masks. The platform could be described as a community of monitoring stations, in where each node is associated with a minimum of ontology on the transmitted data and each variable is related to a metadata definition. Individual nodes can be used only for monitoring or can be integrated with other functions at the local scale, such as home automation systems and other electronic components and could be relocated in future. The described electronic platforms allow monitoring

Conclusioni

Lo sviluppo e connessione di singoli nodi DIY per una campagna di monitoraggio diffusa, per quanto sarebbe consigliabile condurre in futuro test su database più ampi. Inoltre, i dati raccolti mostrano un'elevata sensibilità alle variazioni facendo presagire scenari di analisi delle condizioni di qualità dell'aria e di altre variabili indoor (come nell'esempio) e outdoor basati sui big data. Esempi analoghi sono in fase di sviluppo in numerose altre università e centri di ricerca e le testimonianze riportate sul web e su libri specialistici aumentano rapidamente (es. City Form Lab; Senseable City Lab; Iaac; Media Lab MIT). Google sta implementando numerose applicazioni e funzioni basate sull'analisi dei big data diretti o indiretti con le quali è possibile monitorare la situazione del traffico in tempo reale (si veda Google Earth) o

a large amount of variables that could change over time by modifying sensors, software and hardware solutions. The stations are especially conceived for producing large amount of data in a diffused network in where each node is described by the user on the basis of the list presented in (Tab.1).

Case study. Description, analysis and results

A case study was conducted during the research for simulate a monitoring system based on three stations located in different positions in Turin in indoor environments: a residential building, an office building and a laboratory building (Tab. 2).

The stations connect air-quality sensors and allow in comparing the recorded data (Fig. 1). Three sensors of the MQ series and a temperature and RH sensor (DHT11)

La simulazione di monitoraggio ha dimostrato la semplicità di

sviluppo e connessione di singoli nodi DIY per una campagna di monitoraggio diffusa, per quanto sarebbe consigliabile condurre in futuro test su database più ampi. Inoltre, i dati raccolti mostrano un'elevata sensibilità alle variazioni facendo presagire scenari di analisi delle condizioni di qualità dell'aria e di altre variabili indoor (come nell'esempio) e outdoor basati sui big data. Esempi analoghi sono in fase di sviluppo in numerose altre università e centri di ricerca e le testimonianze riportate sul web e su libri specialistici aumentano rapidamente (es. City Form Lab; Senseable City Lab; Iaac; Media Lab MIT). Google sta implementando numerose applicazioni e funzioni basate sull'analisi dei big data diretti o indiretti con le quali è possibile monitorare la situazione del traffico in tempo reale (si veda Google Earth) o

were employed for each station. The software was conceived for collecting the measurements of different sensors in a nominal variable organized to facilitate the import in a spreadsheet. At the same time, each variables was averaged in the spreadsheets using a laps of sixty seconds, nevertheless this action could be effectively processed adapting the software written for the Arduino board. An objective of the simulation is to demonstrate the possibility of developing monitoring systems based on widespread nodes (low cost and DIY) in order to populate a map in real time thanks to the recorded data (in the case study 5 variables for each station). The visualization is based on the use of Rhinoceros 5, Grasshopper and GHowl, Excel and the shapefile of Turin provided by the City (SIT) (Fig. 2). The script developed in Grasshopper allows increasing the number of con-

costruire la mappa real time della diffusione di malattie e infezioni (es. H1N1, Ginsberg et al., 2009). Anche altre società utilizzano i big data per raggiungere i loro scopi più precisamente e rapidamente rispetto alle modalità di misurazione tradizionale. La Xoom, azienda che si occupa di trasferimenti di denaro, nel 2011 ha potuto identificare e bloccare con azioni mirate la diffusione di carte bancomat clonate nello stato del New Jersey (Mayer-Schönberger & Cukier, 2013; Rosenthal, 2012). I diversi aspetti dei big data real time, modificheranno decisamente il modo di gestire e produrre le modalità di misurazione e analisi delle informazioni. Tuttavia, l'utilizzo dei big data dovrà essere organizzato, in base agli scopi specifici, da gruppi di esperti, costruendo scenari ad hoc. È quindi fondamentale poter disporre di metadati sulle stazioni, e sui singoli dati, tali da essere automaticamente validati.

Nell'insieme delle misurazioni, per quanto concerne i big data, anche quei singoli valori che si scostano dalle aspettative, non modificano la qualità del risultato finale, come invece succede nei monitoraggi scientifici tradizionali, dove si tende all'ottenimento di pochi dati significativi (es. Mayer-Schönberger & Cukier, 2013; Nielsen, 2012; Shirky, 2010). Le tecniche di data-mining e analisi dati si stanno sviluppando in un'ottica di intelligenza collettiva e sapere diffuso (es. Janert, 2011; Segaran, 2007; Larose, 2005) che, grazie al diffondersi di nuovi strumenti open source (Warden, 2011), rendono possibile la predisposizione di nuove modalità di gestione dei database relazionali e dei dataset. L'interesse per soluzioni volte a monitoraggi ambientali DIY è dimostrato dai recenti lavori di Di Justo & Gertz (2013) e Gertz & Di Justo (2012), dallo sviluppo dei progetti di numerose aziende (es. Libelium) e enti pubblici (es. Londa, Parigi, New York).

nected stations, but it requires high computational power of the computer. In addition, the update of monitored data can be managed by using software able to automatically connect serial port and spreadsheet such as PLX-DAQ (Parallax, Inc.). Collected data reveal various phenomena such as the opening of a window, the user profile, the heating profile in the office, the influence on IAQ of an outdoor chimney and trends of vehicular traffic (Q5) near the laboratory (Fig. 3). The temperature line of the residential building shows evening peaks due to the presence of inhabitants and their activities (e.g. cooking food). Differently, the heating system in the laboratory shows a periodic trend on the basis of a predetermined temporal cycle and a set point temperature. Finally, the office temperature results in the combination of a predetermined

on-off cycle of the system and manual adjustments of fan coils. Moreover a different node for transmitting in real time the monitored values was tested using an Ethernet connection. The specific objective of this second experiment is to validate a different modality of data transmission. This node presents three different sensors: a temperature sensor and two gas sensors (CO and TVOC) (Fig. 4). It is possible to read the monitored data via Ethernet using any computer connected to the network. Furthermore, by connecting the USB output of the monitoring node with a computer, it is possible to read, thanks to the developed software, the number of users who are currently connected with the node, the operation of the SD card for storing and information related to the connection (MAC, IP, gateway, and subnet).

Si presume che tale interesse possa aumentare in futuro aprendo ad nuove aree di mercato, focalizzate, ad esempio, sullo sviluppo di ulteriori ricerche considerando la necessità delle PA di produrre (oltre che diffondere) dati ambientali diffusi sul territorio.

REFERENCES

- Barabasi, A.L. (2003), *Link. La scienza delle reti*. Tradotto da B. Antonielli d'Oulx; 1° ed., Perseus Books Group, New York.
- Bar Cohen, Y. (ed.) (2006), *Biomimetics: Biologically Inspired Technologies*, CRC Taylor & Francis Group, Boca Raton.
- Barker, J.A. and Erickson, S. (2005), *Five Regions of the Future: Preparing Your Business for Tomorrow's Technology Revolution*, Penguin Group, New York.
- Braham, W.W. and Hale, J.A. (2007), *Rethinking Technology: a Reader in Architectural Theory*, Routledge, New York.
- Benyus, J.M. (1997), *Biomimicry: Innovation Inspired by Nature*, HarperCollins, New York.
- Cardwell, D.S.L. (1972), *Tecnologia Scienza e Storia*, Tradotto da A. Ca'Rossa; 1° ed., Einemann Educational Publishers, London.
- Celento, D. (2007), "Innovate or Perish. New Technologies and Architecture's Futures", *Harvard Design Magazine*, Vol. 27, pp. 1-9.
- Chiesa, G. (2010), *Biomimetica, tecnologia e innovazione per l'architettura*, Celid, Torino.
- Di Justo, P. and Gertz, E. (2013), *Atmospheric Monitoring with Arduino*, O'Reilly, Sebastopol.
- Ferraris, M. (2003), "Ontologia e oggetti sociali", in Floridi, L. (ed) (2003), *Linee di Ricerca*, SWIF, pp. 269-309, available at: www.swif.it/biblioteca/lr (accessed 17 January 2014)
- Floridi, L. (2013), "THE ONLIFE MANIFESTO", Seminario, Centro Nexta su Internet & Società, Politecnico di Torino, DAUIN. Torino.

Conclusions

The presented case study demonstrates that the development and connection of individual nodes DIY are possible issues for developing a monitoring campaign, even if it is necessary to test it with more complex data sets. In addition, collected data show a high sensitivity to changes. It could be possible, in future studies, to develop scenarios of analysis of air quality and other environmental variables based on bigdata produced by low cost sensors. Comparable examples are being developed in several other universities and research centers and the number of publication related to these topics (e.g. City Form Lab; Senseable City Lab; Iaac; Media Lab MIT) is rapidly growing. Google is implementing a number of applications and functions based on the analysis of direct or indirect bigdata, such as for monitoring

the traffic situation in real time (e.g. Google Earth) or for building a real-time map of the spread of diseases and infections (e.g. H1N1, Ginsberg et al., 2009). Even other companies are using bigdata to achieve more quickly and accurately their goals than with the use of traditional methods of measurement. In 2011, Xoom, a company that deals with money transfers, was able to identify and block illegal actions connected with the spread of cloned ATM cards in the state of New Jersey (Mayer-Schönberger & Cukier, 2013; Rosenthal, 2012). The uses of big data in real time will definitely change the way and methods in which data and information are managed, produced, processed and analyzed. However, as mentioned in this paper, bigdata concerns with several critical points that have to be organized according to any specific monitoring

Gertz, E. and Di Justo, P. (2012), *Environmental Monitoring with Arduino*, O'Reilly, Sebastopol.

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S. and Brilliant, L. (2009), "Detecting influenza epidemics using search engine query data", *Nature*, No. 457, pp. 1012-1015.

Hashimoto S. and Dijkstra R. (2004), "Chip City", *Verb Connection*, No. 3, p.46-53.

Hochberg, L.R., Bacher, D., Jarosiewicz, B., Masse, N.Y., Simeral, J.D., Vogel, J., Haddadin, S., Liu, J., Cash, S.S., van der Smagt, P. and Donoghue, J.P. (2012), "Reach and grasp by people with tetraplegia using a neurally controlled robotic arm", *Nature*, No. 485, p. 372-375.

Janert, P.K. (2011), *Data Analysis with Open Source Tools*, O'Reilly, Sebastopol.

Larose, D.T. (2005), *Discovering Knowledge in Data. An Introduction to Data Mining*, John Wiley & Sons, Hoboken.

Maldonado, T. (1992), *Reale e virtuale*, Feltrinelli, Milano.

Mayer-Schönberger, V. and Cukier, K. (2013), *Big Data. Una rivoluzione che trasformerà il nostro modo di vivere e già minaccia la nostra libertà*, Garzanti, Milano.

Milgram, P. & Colquhoun, H. (1999), "Chapter 1: A Taxonomy of Real and Virtual World Display Integration", in Ohta, Y. & Tamura, H. (eds.), *Mixed reality: merging real and virtual worlds*, Springer, New York, pp. 5-30.

Mitchell, W.J. (2005), "Construction Complexity", in Martens, B. and Brown, A. (eds.), *Computer Aided Architectural Design Futures 2005*, Springer, Netherlands, pp. 41-50.

Negroponte, N. (1995), *Essere digitali*, Sperling & Kupfer Editori, Milano.

Nielsen, M. (2012), *Le nuove vie della scoperta scientifica. Come l'intelligenza collettiva sta cambiando la scienza*, Einaudi, Torino.

Odifreddi, P. (1994), "Epistemologia e Ontologia virtuali", *Cenacolo Interdipartimentale di Torino*, 27 Maggio 1994, Torino.

purpose by groups of experts. It is essential to produce metadata regarding the monitoring stations, and methods for automatically validate data and dataset. It is important to remember that, in bigdata measurements, those individual values that deviate from expectations do not change the quality of the final result, as it happens in traditional scientific monitoring based on a few relevant data (Mayer-Schönberger & Cukier, 2013; Nielsen, 2012; Shirky, 2010). New techniques of data mining and data analysis are under developing in order to improve the use of collective intelligence and diffuse knowledge (Janert, 2011; Segaran, 2007; Larose, 2005) and, thanks to the spread of new open source tools (Warden, 2011), it possible the establishment of new methods of management of database and relational datasets.

The effectiveness of DIY solutions to

environmental monitoring is, also, demonstrated by the recent works of Di Justo & Gertz (2013) and Gertz & Di Justo (2012) and other projects and researches of many companies (e.g. Libelium) and public institutions (e.g. London, Paris, New York). This interest will increase in the future and could be translated into new market areas, focusing, for example, on the development of further researches and instruments considering the needs of the PA to produce (and spread) environmental data covering the entire territory.

Ocelli S. and Staricco L. (2002), *Nuove tecnologie di informazione e di comunicazione e la città: Elementi di riflessione*, Franco Angeli, Milano.

Oxman, R. (2006), "Theory and design in the first digital age", *Design Studies*, Vol. 27, No. 3, p. 229-265.

Rivoltella, P.C. (ed) (2010), *Ontologia della comunicazione educative. Metodo, ricerca, formazione*, Vita e pensiero, Milano.

Rosenthal, J. (2012), "Special Report: International Banking. Big Data Crunching the numbers", *The Economist*, 19 may 2012, pp. 7-8, available at: <http://www.economist.com/node/21554743> (accessed 04 January 2014)

Rothenberg, J. (1989), "The nature of modelling", in Widman, L.E., Loparo, K.A. and Nielsen, N.R., *AI, Simulation & Modelling*, John Wiley & Sons, New York, pp. 75-92.

Sakamoto, T., Ferré, A. et al. (eds.) (2008), *From Control to Design. Parametric/Algorithmic Architecture*, Actar, Barcellona.

Sass, L. and Oxman, R. (2006), "Materializing design: the implications of rapid prototyping in digital design", *Design Studies*, Vol. 27, pp. 325-355.

Segaran, T. (2007), *Programming Collective Intelligence*, O'Reilly, Sebastopol.

Shirky, C. (2010), *Cognitive Surplus: Creativity and Generosity in a Connected Age*, Penguin Group, London.

Warden, P. (2011), *Big Data Glossary*, O'Reilly, Sebastopol.

Weinberger, D. (2012), *Too Big To Know: Rethinking Knowledge Now That the Facts Aren't the Facts, Experts Are Everywhere, and the Smartest Person in the Room is the Room*, Basic Books, New York.

Wiener, N. (1950), *Introduzione alla cibernetica: l'uso umano degli esseri umani*, Tradotto da D. Persiani; 1° ed. Houghton Mifflin Company, Boston.

Xu, X. (2012), "From cloud computing to cloud manufacturing", *Robotics and Computer-Integrated Manufacturing*, Vol. 28, pp. 75-86.

City Form Lab: <http://cityform.mit.edu/> (accessed July 2013)

The Onlife Initiative: <http://ec.europa.eu/digital-agenda/en/onlife-initiative> (accessed July 2013)

Senseable City Lab: <http://senseable.mit.edu/> (accessed September 2014)

Media Lab: <http://www.media.mit.edu/> (accessed September 2014)

Iaac: <http://www.iaac.net> (accessed September 2014)