Roberto Minerva
Telecom Sud Paris

The quest for data and the virtual continuum
the Quest for Data and the Virtual Continuum

Roberto Minerva
Telecom SudParis
Institut Mines-Telecom
In the Programmable World, All Our Objects Will Act as One

Wired - BILL WASIK 05.14.13. available at:
https://www.wired.com/2013/05/internet-of-things-2/

We are talking about a gigantic network of all things (millions of systems). It has to be:

- Open
- Programmable (API based)
- Secure
- Accountable
- Manageable
- Reliable
- Always available
- Usable
- Democratic
- Inclusive
- Supporting several stakeholders
- Enabling different business models
Application Domains and the Quest For Data

https://datavisualization.ch/wp-content/uploads/2010/05/cph_wheel_04.png
IoT: Application Domains

In each of these application domains, data will have a fundamental importance.
Example: How do we understand and reason about Cities?

THE CITY, TECHNOLOGY, AND HISTORY*

PAUL MEADOWS

University of Nebraska

THE APPROACH THROUGH INTRA-URBANISM

Over two decades ago a leading United States urban sociologist, Professor Niles Carpenter, opened a discussion of urban sociology with this statement: “Recent trends in the field of sociology should be epitomized in a four-word phrase — the quest for data.” In retrospect, one might still sense the importance of this empirical bent, still ask the elementary question, data about what? So far as urban sociology is concerned, it is perfectly obvious that so long as it is data about some relationship concerning social life within the American city — whether trend, stage, cause-effect, fact-implication, problem-policy — which is to be discovered, nothing else has ever seemed to count. Urban sociology has been and is yet literally (and without reservation apparently) the sociology of life within the city.

This approach to urban sociology, which we may designate as the sociology of intra-urbanism because of the manner in which such phenomena are interpreted solely in terms of the city itself, has been characterized by both purely intellectual as well as markedly pragmatic interests. As an intellectual curiosity, urban sociology represents the emergence of the city as in itself a legitimate object of sociological study. The city is sui generis; hence, the sociology of city life. This perspective was proclaimed in an extraordinarily influential volume of papers published by the University of Chicago Press in 1924: The City, edited by R. E. Park, E. W. Burgess, and R. D. McKenzie. The theoretical position taken by these authors is indicated in the initial paper by Professor Park: “The City: Suggestions for the Investigation of Human Behavior in the Urban Environment.” The subsequent ecological, personality, and institutional investigations of a generation of urban sociologists are foreshadowed in some of the other papers in this volume: McKenzie’s “The Ecological Approach to the Study of the Human Community”; Park’s “The Mind of the Hobo,” and his famous paper on the metropolitan daily newspaper. Since then, the classroom texts in urban sociology3 follow rather closely this thematic organization of this field. The much later, masterful essay by Professor Louis Wirth, summarizing and organizing the theory of a sociology devoted to the study of intra-urbanism and significantly titled “Urbanism as a Way of Life,” has been one of the

* Read at the Seventh Annual Congress of Sociology of the Mexican Association of Sociology, December 3-7, 1956, Universidad de Nuevo Leon, Monterrey, Mexico.


3 American Journal of Sociology, 44 (July 1938), pp. 1-25.
A check list for a Measurable Environment

- What do you want to measure
- How to measure your features
- What sensing technologies
- How to deploy enough sensors to have a valid coverage and measurement of your features
- What communications means
- How many data you have to handle
- What management processes
Measurable City: example counting people without infringing Privacy

The Future Centre in Venice is working at the monitoring in quasi real-time on the pedestrian flows in the city center. The goal is to measure the pedestrian traffic and keeping the anonymity and privacy of users. The project has been using low-cost sensors and devices (50-100 euros) with a small size (two cigarette packs) in order to acquire video flows of passing by people and to process it locally without any leak or privacy violation. These devices will provide their Id, the time and the number of people that have been detected. Ideally these objects could be scattered in many places of the city and freely transit their data (e.g., through twitter). In such a way, interested developers could crate new applications based on these data..
Towards a Virtual Continuum, an enabler for the Programmable World
Evolutionary roadmap for key functional elements

**M2M**
- Key elements:
  - Separated applications
  - Ad-hoc designed modules
  - Ubiquitous connectivity
  - SIM management
  - International agreements
  - Embedded SIM

**IoT**
- Key elements:
  - Low-cost standard sensors
  - Short range communication
  - Capillary and macro netw.
  - Horizontal Services
  - Data aggregation in cloud
  - Third party development

**Virtual Continuum**
- Key elements:
  - “Virtual Objects”
  - Mirroring Things in cloud
  - Object Semantics
  - Data integration, federation and portability
  - Cloud as developing platf.
A different approach to data

- From Wikipedia: Spime is a neologism for a futuristic object, characteristic to the Internet of Things, that can be tracked through space and time throughout its lifetime. They are essentially virtual master objects that can, at various times, have physical incarnations of themselves. An object can be considered a spime when all of its essential information is stored in the cloud. Bruce Sterling sees spimes as coming through the convergence of six emerging technologies, related to both the manufacturing process for consumer goods, and through identification and location technologies. Depending on context, the term "spime" can refer to both—the archetype, as designed by the developer, or a user-specific instance of it.

- Data representing an object is then a set of triplets

\[(\text{value} + \text{location} + \text{time})\]

- Plenty of novel applications become immediately possible

- SPIME is data in space and time!
The Virtual Continuum (\(=\) SPIME + APIs)

A virtual environment is a software feature that allows customers to use an entire (virtualized) computational and communication environment tailored to their specific needs.

The Virtual Continuum is the constant entanglement between real objects and their representations in the network. Events, actions, data on a physical object will be represented in the virtual world and vice versa. The Virtual Continuum makes possible the close relation between atoms and bit.
Each Virtual Object should have

- A unique identifier
- A set of states and data representing it in time and space
- Communication mechanisms
- A means (a protocol or API) to synch with the real object
- A contract (managing ownership and authorization to use)
- An API for augmentation of functionalities
- Some logic to control the “entanglement”

Communication, storage and processing are strong requirements for the Virtual Object (or for at least one digital twin)
Identity and Status of an Object ...

- **Basic Object Attributes** (what type of sensor, what measures, ...)
  - Object Id
  - Object Status
  - Object Location
  - Object Time - Period

- **Authentication and Authorization Attributes**
Virtual Continuum is Virtualization of Resources

ICT Resources

- Local Services
- Interface

Virtual Continuum

- Global Services
- Extended Functions/Interface
- API

- Strong Entanglement

- Each Resource is representable
- Each resource is programmable
- Each Resource can be functionally augmented
- This concept could also be applied to logical objects
- Virtual Objects can represent past states and often “intended” future states

Physical/Logical Resources

- Physical Resource
- Interface

- Weak Entanglement
The Context for Virtual Continuum

Computing Cloud

images of terminals and objects (Proxying and Security Functions)

applications & services

Virtual Environment

Brokering and Aggregation Function

Virtual Environment

Secure, always best connected virtualized channel

physical world

mobile devices

smart objects

network resources

processing resources

storage resources

Virtualized resources

Secure, always best connected virtualized channel

Computing Cloud

Virtualized resources
An Example: the virtualized home

- How Many Sensors!! And what a complicated system
- What useful Services?
- How to use them?
- How to deploy them
- How to maintain them?
An Example: Terminal Virtualization

iPhoneX is so powerful: if you delete one of your contacts, that person will die in the real life.
Cloud Augmentation of the Terminal and Smart Objects in general
What are Internet-Connected Things?
Active/passive, with/without context

<table>
<thead>
<tr>
<th>Object Type</th>
<th>Generic Info</th>
<th>Contextualized Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passive Objects</td>
<td>A Tag, A Pointer to some information</td>
<td>Info + a location</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reactive Objects</td>
<td>A switch at home (turn it on/off), A smart meter</td>
<td>Home Automation (when temperature reaches 20 C stop heating)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autonomous Objects</td>
<td>A Vending Machine, An Intelligent Fridge</td>
<td>A Cleaning Robot</td>
</tr>
</tbody>
</table>
How Many Things?

THE INTERNET OF THINGS
AN EXPLOSION OF CONNECTED POSSIBILITY

How much Data (and traffic)?

Table 3. Summary of Per-Device Usage Growth, MB per Month

<table>
<thead>
<tr>
<th>Device Type</th>
<th>2014</th>
<th>2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonsmartphone</td>
<td>22 MB/month</td>
<td>105 MB/month</td>
</tr>
<tr>
<td>Wearable Device</td>
<td>141 MB/month</td>
<td>479 MB/month</td>
</tr>
<tr>
<td>Smartphone</td>
<td>810 MB/month</td>
<td>3,081 MB/month</td>
</tr>
<tr>
<td>4G Smartphone</td>
<td>2,000 MB/month</td>
<td>5,458 MB/month</td>
</tr>
<tr>
<td>Tablet</td>
<td>2,076 MB/month</td>
<td>10,767 MB/month</td>
</tr>
<tr>
<td>4G Tablet</td>
<td>2,913 MB/month</td>
<td>12,314 MB/month</td>
</tr>
<tr>
<td>Laptop</td>
<td>2,641 MB/month</td>
<td>5,589 MB/month</td>
</tr>
<tr>
<td><strong>M2M Module</strong></td>
<td><strong>70 MB/month</strong></td>
<td><strong>368 MB/month</strong></td>
</tr>
</tbody>
</table>

- = 2.33 MB /Day
- = 12.2 MB /Day

Source: Cisco VNI Mobile, 2015

## Bytes per Object (Sensor) per Day

### Message length in Bytes

<table>
<thead>
<tr>
<th>Messages per day</th>
<th>1</th>
<th>8</th>
<th>64</th>
<th>128</th>
<th>256</th>
<th>512</th>
<th>1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 per day</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>64</td>
<td>128</td>
<td>256</td>
<td>512</td>
</tr>
<tr>
<td>1 per H</td>
<td>24</td>
<td>24</td>
<td>192</td>
<td>1536</td>
<td>3072</td>
<td>6144</td>
<td>12288</td>
</tr>
<tr>
<td>1 per min</td>
<td>1440</td>
<td>1440</td>
<td>11520</td>
<td>92160</td>
<td>184320</td>
<td>368640</td>
<td>737280</td>
</tr>
<tr>
<td>1 per sec</td>
<td>86400</td>
<td>86400</td>
<td>691200</td>
<td>5529600</td>
<td>11059200</td>
<td>22118400</td>
<td>44236800</td>
</tr>
<tr>
<td>1 per 500 ms</td>
<td>172800</td>
<td>172800</td>
<td>1382400</td>
<td>11059200</td>
<td>22118400</td>
<td>44236800</td>
<td>88473600</td>
</tr>
<tr>
<td>1 per ms</td>
<td>86400000</td>
<td>86400000</td>
<td>691200000</td>
<td>5529600000</td>
<td>11059200000</td>
<td>22118400000</td>
<td>44236800000</td>
</tr>
</tbody>
</table>

### Smartphone average data consumption

<table>
<thead>
<tr>
<th></th>
<th>in Mbytes</th>
<th>in Mbps</th>
<th>Peer to peer</th>
<th>Video Streaming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Western Europe</td>
<td>86.4</td>
<td>0.008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>691.2</td>
<td>0.064</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5529.6</td>
<td>0.512</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>11059.2</td>
<td>1.024</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>22118.4</td>
<td>2.048</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>44236.8</td>
<td>4.096</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>86400</td>
<td>8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Source
A few Considerations

- **Data Sources**
  - Sensor data rate generation strongly depends on the sensor and the specific application. When the number of sensors grows, the aggregate data generation rate grows.
  - Data load increases with the size of the payload (some sensors can generate small amounts of data with a high rate).
  - Multimedia sensors (e.g., cameras) tend to generate large streams of data.
  - When the Dt between two messages $\rightarrow 0$, then the sensor generates a stream of data.
  - Streams of sensor data can exhibit similar patterns as multimedia data, but sometimes they have different requirements.
  - Number of "streaming IoT devices" can have a huge impact on the network.

- **Mediation**
  - When many sensors are mediated by a Gateway, then:
    - Messages may be processed and reduced in quantity (aggregation).
    - The rate of outgoing messages from GW to Server may be high and may generate a stream.

- **Sensor behaviour**
  - Some IoT systems exhibit impulsive behaviours:
    - A Container Ship with M2M devices entering a harbor generates a spike of signalling and data.
    - An alarm generated by several sensors in a specific area.
    - Malfunctioning: a sensor that generates many messages and stresses nearby resources.

- **Data Collection (data sinks)**
  - Data generated by sensors may be stored and processed locally (edge computing) or transferred as raw data over the network.
  - Data sets of raw data can be very useful to study particular phenomena, but transferring them to cloud systems could be expensive.
IoT Devices with Cellular

- Ericsson predicts 1.5 billion IoT devices with cellular subscriptions by 2021

- Does it make sense to postulate
  - 0.5 Billion of individual Sensors directly connected to Cellular Net
  - 1 Billion of gateways directly connected
## The “IoT Data Avalanche”

<table>
<thead>
<tr>
<th>CASE</th>
<th>1 mes per day length 1 byte</th>
<th>1 mes per min length 64</th>
<th>1 mes per 0.5 sec length 8 byte</th>
<th>1 mes per ms length 1000 byte</th>
<th>1 message per sec 1000 bytes</th>
<th>1 message per sec 256 bytes</th>
</tr>
</thead>
<tbody>
<tr>
<td>CASE 1</td>
<td>0,00000001</td>
<td>9,216</td>
<td>1,3824</td>
<td>864</td>
<td>0,864</td>
<td>0,221184</td>
</tr>
<tr>
<td>CASE 2</td>
<td>0,009216</td>
<td>92,16</td>
<td>138,24</td>
<td>8640</td>
<td>8,64</td>
<td>2,21184</td>
</tr>
<tr>
<td>CASE 3</td>
<td>0,013824</td>
<td>1,3824</td>
<td>138,24</td>
<td>864000</td>
<td>86,4</td>
<td>22,1184</td>
</tr>
<tr>
<td>CASE 4</td>
<td>864</td>
<td>8640</td>
<td>864000</td>
<td>86400000</td>
<td>86400000</td>
<td>663552000</td>
</tr>
<tr>
<td>CASE 5</td>
<td>0,864</td>
<td>8,64</td>
<td>86,4</td>
<td>8640</td>
<td>86400000</td>
<td>2592000000</td>
</tr>
<tr>
<td>CASE 6</td>
<td>0,221184</td>
<td>2,21184</td>
<td>22,1184</td>
<td>221184</td>
<td>221184</td>
<td>22118400</td>
</tr>
</tbody>
</table>
The IoT Data Avalanche (cont.)

<table>
<thead>
<tr>
<th>1 B connected Objects</th>
<th>Pbytes/day</th>
<th>Pbytes/Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 mes per day length 1 byte</td>
<td>0,000001</td>
<td>0,00003</td>
</tr>
<tr>
<td>1 mes per min length 64 bytes</td>
<td>0,09216</td>
<td>2,7648</td>
</tr>
<tr>
<td>1 mes per 0.5 sec length 8 bytes</td>
<td>1,3824</td>
<td>41,472</td>
</tr>
<tr>
<td>1 mes per ms length 1000 bytes</td>
<td>86400</td>
<td>2592000</td>
</tr>
<tr>
<td>1 message per sec 1000 bytes</td>
<td>86,4</td>
<td>2592</td>
</tr>
<tr>
<td>1 message per sec 256 bytes</td>
<td>22,1184</td>
<td>663,552</td>
</tr>
</tbody>
</table>

Total Mobile Traffic in Ebytes per month in 2015
Source: Ericsson

| Western Europe | 0,8 |
| USA           | 1,3 |
| World Total   | 5,4 | 5400 |
IoT Data and ... Identity of Things

Things have Identities (and Owners)

“Me”  Smart Thing

Sensors

People have Identities and use Things

Me

Service Provider

Personal Profiling

Who, Where, When, What, Why, ...

Functional Relations
(events and commands)
Aggregating Data per Identity ...

- Who is the Owner of all these Data?
- Who has the right to extract info?

“OUR” Smart Things

Raw data to be transformed into Info

Functional Profiling

Personal Profiling

Who, Where, When, What, Why, ...

Events and commands

Bigger DATA

50 B Devices

~ 2MB/day

(Average Aggregated Traffic of M2M Devices)

= ~ 88.81 petabytes /day
How Smart Objects will communicate

Network Intelligence (e.g., IMS) is a hierarchical model based on the assumption that control has to be exerted by a few specialized control nodes.

Client – Server model disregards the network aspects and can lead to a tragedy of commons (misuse of common networking resources).

Other mechanisms (message based like PubSub) can be more attractive for IoT.

This is a reason for different IoT protocols.

… Is it there a single communication paradigm for IoT?
Van Jacobson (on a google video): 90% of the data traffic is not related to communications services, but to data retrieval

A new data network architecture is needed

Content Centric Networking
A Self-Organizing Network That Meets Information Needs

What Is It? A new approach to networking that enables networks to self-organize and push relevant content where needed.

Content-centric networking enables communication to happen anywhere, anytime, and with any device - using any available means.

Information centric networks
Source: http://www.surrey.ac.uk/ics/research/cognitivenetworks/
5G and PubSub

- Intelligent Routing of Events and Messages thanks to SDN
- Transaction Management
- R.T. extraction of Knowledge

We need to bring Intelligence at the Edge of the Network
A Middleware View of the Virtual Continuum
Business Challenges
Servitization is the capability of creating a link between a (physical) product and a set of services and enriched functionalities that extend, complement, and add value to the product itself.
A long value chain opens up opportunities for many Actors

The Ecosystem Challenge

- **Module/modem supply**
  - SIM cards
  - Sensors
  - Actors
  - Aggregators
  - Transponder

- **Smart object**
  - Vending machines
  - PNDs
  - Cars
  - Cameras
  - Computer

- **Network operator**
  - Network
  - Connectivity
  - Availability
  - Quality

- **Service enabler**
  - Platform
  - Enabling capabilities (e.g., QoS)
  - Applications

- **System integrator**
  - Interfaces
  - Solution build-up
  - Hardware
  - Backend

- **Service provider**
  - Packaging/Bundling
  - Service Provisioning
  - CRM
  - Billing

- **Reseller (Bus. cust.)**
  - Uses service
  - Resells services

- **Customer**
  - Buys service
  - Uses service

**Average Value Share Distribution**

- 5 – 10%
- 15 – 20%
- 30 – 40%
- 15 – 20%
- 10 – 20%

Value highly distributed across the Value Chain – creating the opportunity for a consolidator

Source: Nokia Siemens Networks
An Example: Cultural Heritage

Scala Contarini del Bovolo, Venice

Let’s Virtualize it

- We need to capture data describing the object that span over time …
  - Understanding the past and
  - Framing (history and data) in the current world
- See Venice Time Machine from EPFL
  - Collection of all historical data and then creation of linkage between this huge archive of histories
  - [https://www.youtube.com/watch?v=QTBkuyFblzO](https://www.youtube.com/watch?v=QTBkuyFblzO)
- We need to tell a compelling story
  - Data collection about the object, useful for historians but also useful to citizens and visitors
  - Extracting data not only from documents but also from other «information» sources (e.g., oral tradition and so forth)
  - A similar approach applicable to buildings, artifacts, oral tradition, music
- This means …
Transforming it into a SPIME

Data representing an historical object could then be a set of triplets

(location, time, value)

Object relationships with other objects tells a story

Plenty of data to mine

Plenty of (virtual reality) applications become immediately possible
Time Window Weimar: Students Map their Town's History through Augmented Reality
An Example: Immersive presence

- Fulvio Dominici started [www.ultramundum.org](http://www.ultramundum.org) with the goal of providing new solutions for the immersive presence.
- It is based on an intuition: adding the TIME dimension to the 3D Virtual World
  - Possibility to surf and experience environments as they are now or back in time
  - Creation of personalized Worlds
- Users will spend more time in virtual worlds than in real worlds
- Practical Applications:
  - Cartography for Turin Olympic games
Build your own APP with NEARBY function of Wikipedia

Many objects in Wikipedia have spatial coordinates.

There is a «Nearby search function»

https://en.m.wikipedia.org/wiki/Special:Nearby

And an API definition:
http://blog.wikimedia.org/2013/01/31/geodata-a-new-age-of-geotagging-on-wikipedia/
An ICT Architecture for Cultural Heritage
Five dimensions of Cultural Heritage

- **First Dimension: Identification and preparation of the Cultural Heritage Content**
  - These activities are cross-disciplinary: they aim at identify the Cultural Content. The existing sources for data identified (e.g., Museums, Historical Archives, ...) in addition hypothesis for relating the different sources should be put forward. In addition the means to be used to create new data or collect new information should be stated and elaborated (e.g., how to collect the oral tradition).

- **Second Dimension: digitalization of the Cultural Heritage Content**
  - The identified sources should be progressively digitalized (it can take a long time) without impacting too much on the original material (e.g., old books) and a «big data» like infrastructure put in place. At this stage also discovery of new relationships and «automatic reasoning» on available data should be elaborated and performed.

- **Third Dimension: easing the access to the (digitalized) Cultural Content**
  - At this stage, mechanisms for easing the access to available (original or inferred) information should be studied and made available to user. Extensive research on user interaction, recommendation engines and new forms of data representation should be conducted.

- **Fourth Dimension: the Ecosystem View on the Cultural Heritage**
  - Integration of Measurable City data with Cultural Heritage ones. This wealth of data should be made available to enterprises, users and organization in order to promote the city or the territory and to create a strong link between the cultural heritage and the enterprises operating in the tourism or cultural industries.

- **Fifth Dimension: the creation of new Cultural Content**
  - New Digital technologies can be made available to artists, designers and also users in order to create new (digital) content. Examples could be the Art Lab (as an implementation of the FabLabs for art), new forms of TV, etc.
An Architecture for Digital Humanities

(a part of)

Feeds from the Measurable City
Grazie

Roberto TSP
roberto.minerva@telecom-sudparis.eu
or
roberto.minerva@gmail.com