Not So Volunteered Geographic Information

Fabien Girardin
International Citizen Cartography Meeting
Gijón, Spain. July 30, 2009

Lab.

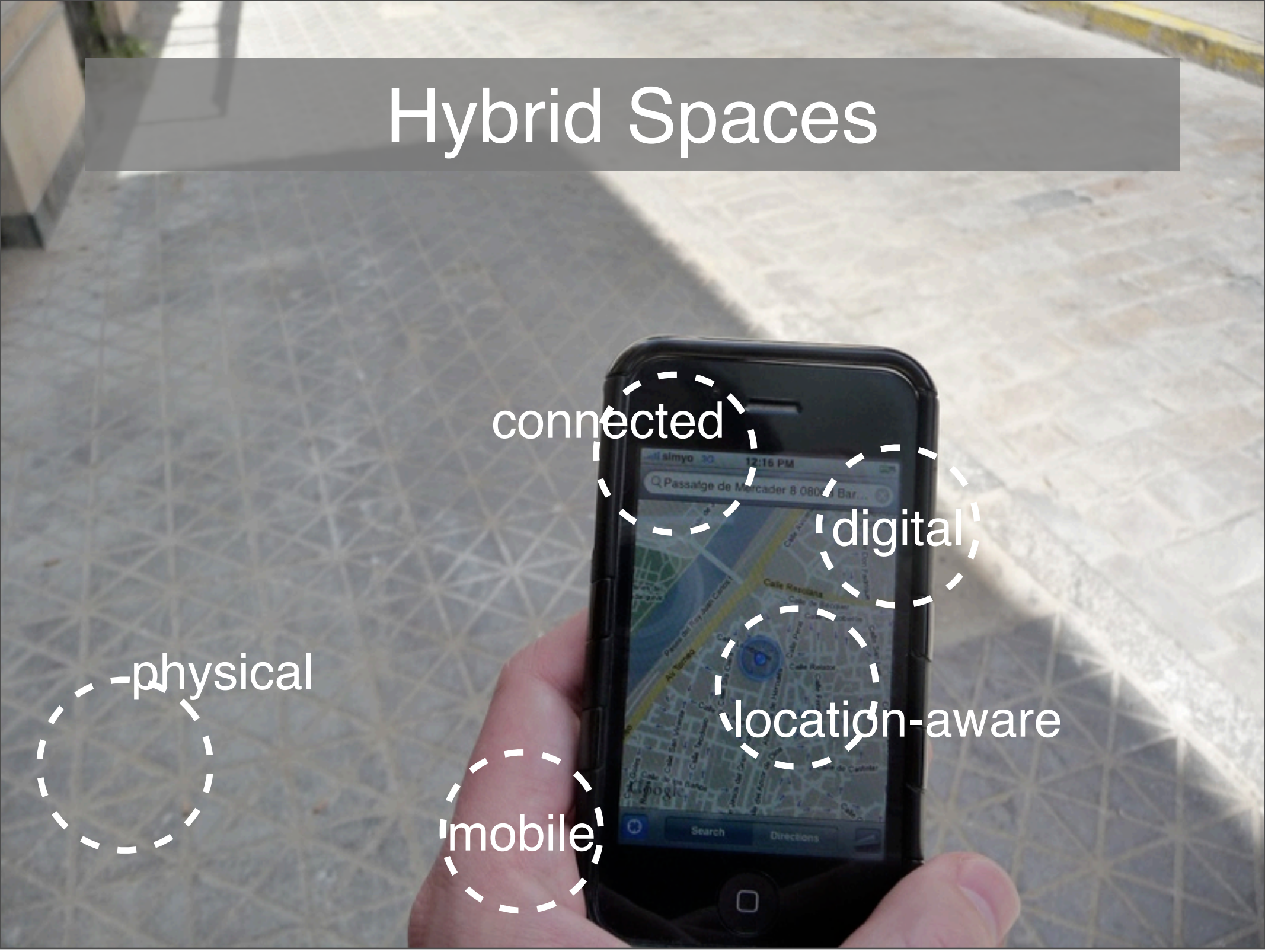
My name is Fabien Girardin, I am a research and engineer at Lift lab a research agency that tries to understand, foresee and build upon the socio-technical evolutions. In this talk, I will particularly discuss the research I have been leading in the domain of urban informatics. Most of the works I will present are fruits of my PhD thesis that I have recently defended.

“Current technology requires information to be served from somewhere and delivered to somewhere [...] ***at geographic scales a bit always has an associated location in real geographic space****.”*

Goodchild, M. F. (1997). Towards a geography of geographic information in a digital world. Computers, environment and urban systems, 25(6):377–391.

This unique quote from Michael Goodchild strongly links me to this meeting on Citizen Cartography. Every communicated bit has one or many geographic reference attached to it (origin, destination, route, ...).

Hybrid Spaces



This presence of bits has become ubiquitous with the emergence of mobile, wireless and location-aware technologies, mixing and sharing locations in the real geographic space (physical space) and its digital counter part.

Implicit interactions

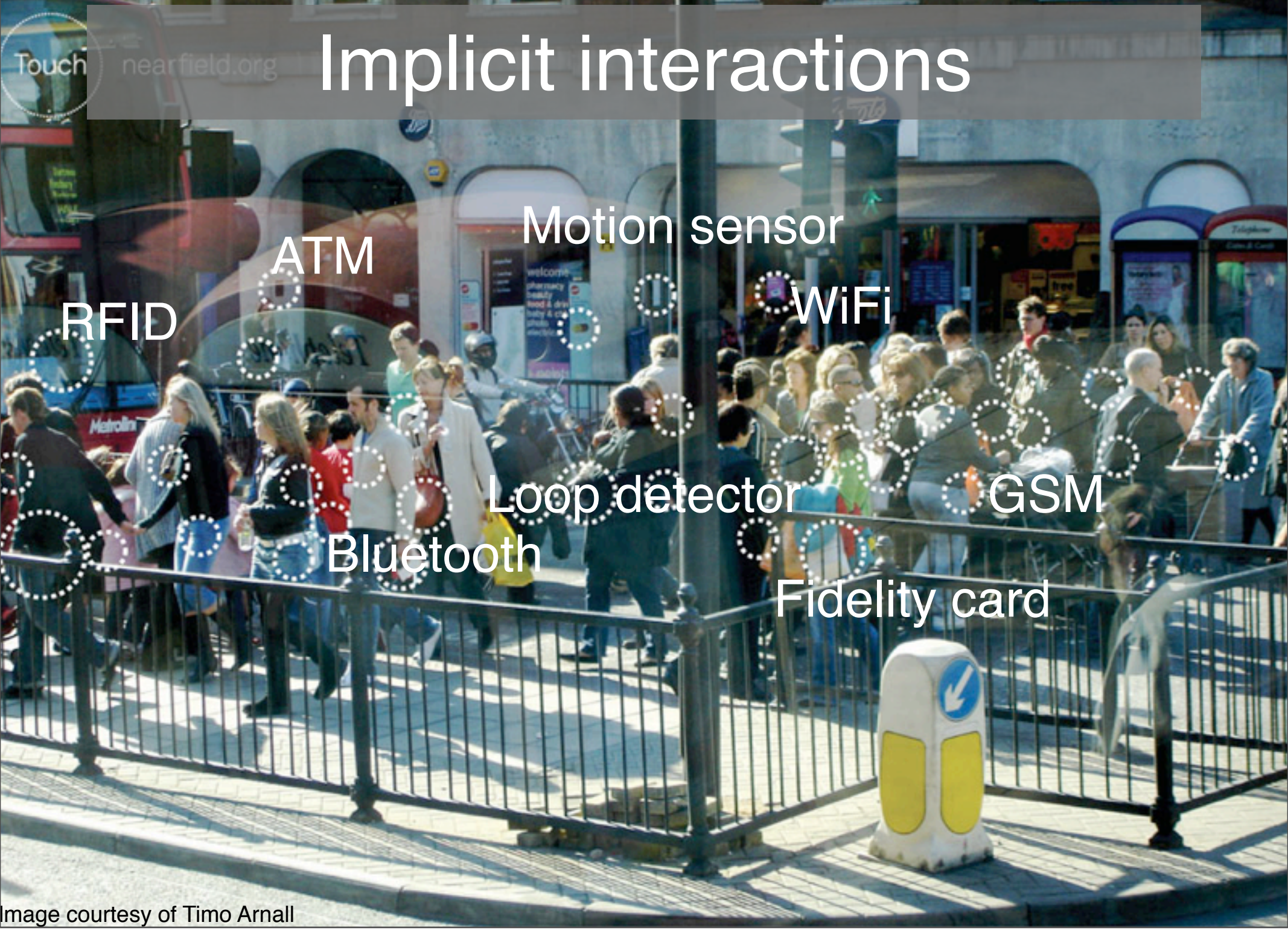


Image courtesy of Timo Arnall

This ubiquitous presence of bits forms an informational layer that were are most of time not aware of. Indeed, communications take place without our awareness through our implicit interactions with soft urban infrastructures such as an RFID card to enter the bus, a fidelity card (with barcode) to get a discount, your mobile phone that performs a location update operation with the network of your mobile operator, a loop detector that counts traffic, a motion sensor that opens an automatic door, your bluetooth enabled phone that scans its environment, your Blackberry that automatically connects to a public WiFi access point and so on...

"Not So" Volunteer Geographic Information

- Digital shadows, footprint, fingerprint, breadcrums are now essential to our lives and the contemporary city functions
- Individual, anonymous, aggregated
- Levels of geographic granurality (location cloacking)
- Private and public sources (current trend to open the data)

The logs of these implicit interactions represent an amount "Not So" Volunteer Generated Geographic Information (some call them digital shadows, footprints, breadcrums, ...) that are essential to our lives and the conemporary city functions (e.g. a GSM network depends on mobile phones location updates, loop detectors help model the flows). These information take multiple forms: they can represent individuals, ananymous individuals, flocks; with fluctuating quality depending on the granularity and the potential location cloacking algorithm. Some are in the public domain , but a majority are kept in a closed circuit, unaccessbile to the citizen who actually produce them.

The ubiquitous technologies that afford us new flexibility in conducting our daily activities are simultaneously providing the means to study our activities in time and space.

In other words, the ubiquitous technologies that are becoming an essential part of our daily activities afford use new means to map our activities and use of the space.

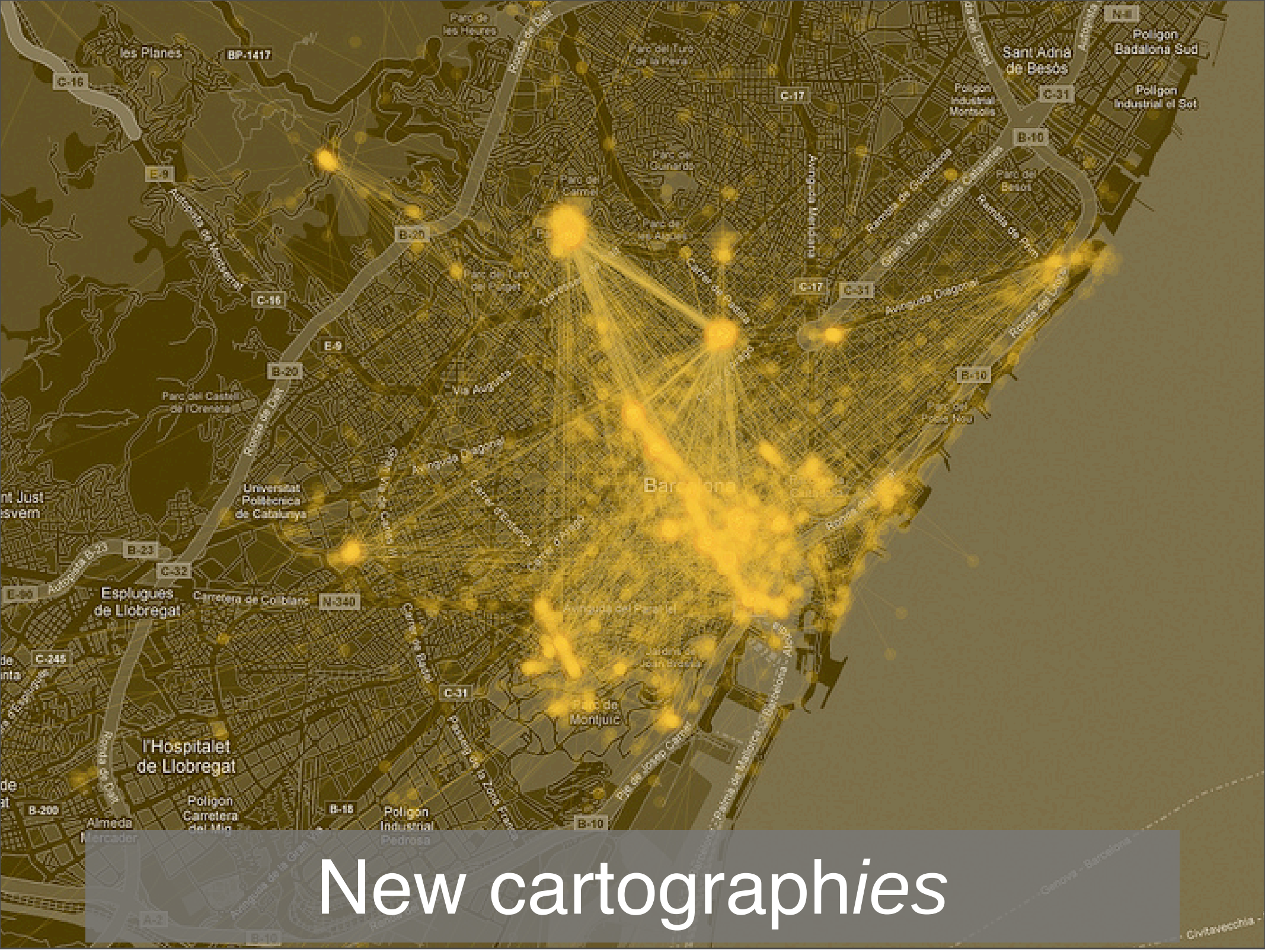
Opportunities

This presence of information later creates new opportunities.

“When many individual diagrams are aggregated to the level of cities and regions, these visualizations may provide geographers, for the first time, with truly dynamic maps of dynamic human processes. ***One might imagine them as twenty-first century “weather maps” of social processes.”***

Zook, M., Dodge, M., Aoyama, Y., and Townsend, A. (2004). New digital geographies: Information, communication, and place. Geography and Technology, pages 155–176.

Opportunities that were imagined not long time ago... (“weather maps” of social processes”)



New cartographies

.. have become reality, through the development of new “cartographies”, that is new practices of drawing map. These new cartophgraphies become means reveal previously invisible urban processes.

Sentient and responsive

As technologies are getting closer to us, we also have new means to:

- **Observe and improve** rather than predict and accommodate
- **Understand their implications** (co-evolution)
- **Learn from design errors** (evaluation)
- **Offer appropriate functions**

This ability to reveal the invisible (sentient) can help us in moving from predicting and accomodating urban spaces, infrastructures or services to actually observe how they function (learn from the design errors), improve them (understand their implications) and eventually offer appropriate services (responsive).

Privacy and ethical issues

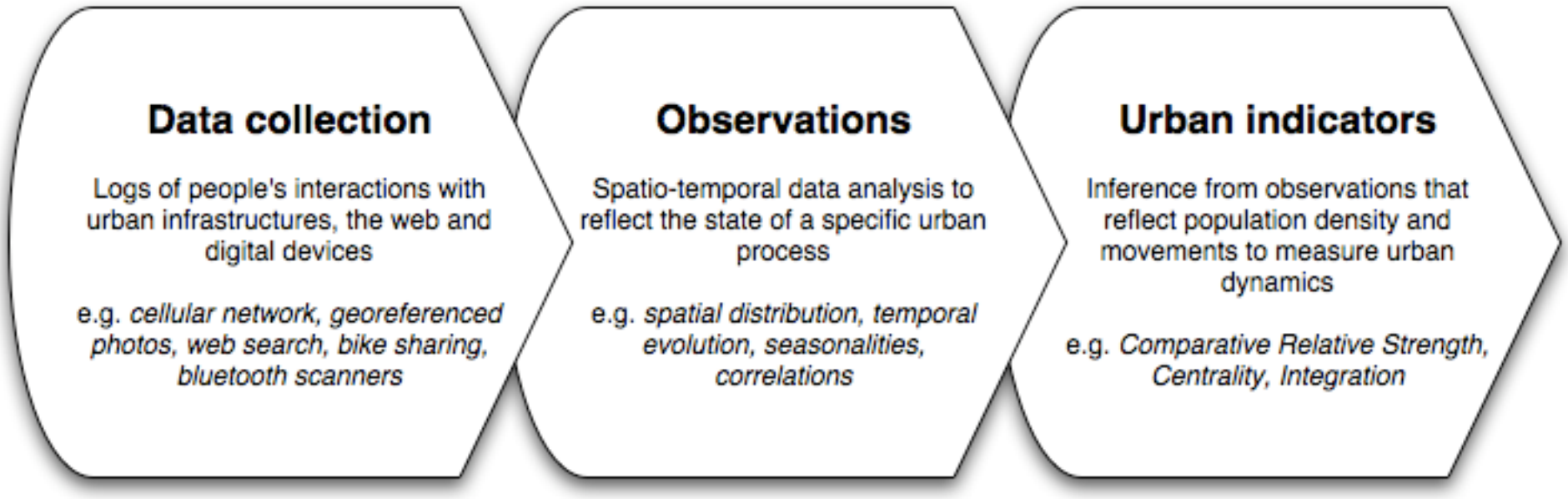
How much are we willing to give to get a service in return?

Gathering data from people without their knowledge? Who owns the data?

Myth of the perfect power: Panopticon vs oligopticons

This opportunity of sentient and responsive cities do not come without creating social and ethical issues. My research participates to the debat in creating positive value to these data. Sof infrastructures rely on digital footprints to perform and offer the basic functions of a contemporary city. So the question as not become whether we want this data or not, but what do we do with them. As a society, we will need to chose how much we are willing to give to get a service in return? The success of fidelity cards in chains of supermarkets is an example of the trade-off privacy vs. value. The debate must also consider who owns the digital footprints produced in hybrid spaces (the owner of the physical space or the owner of the infrastructure, or the service providers, or the actual individual who produced the log, or ?). However, this kind of debate might be mislead with the myth of the perfect power, contradictory with the the messiness and unpredictability of the world. So instead of fearing the emergence of a digital panopticon, it seems were are producing oligopticons that sees very little information very good. (Latour: blind but plugged in, partially intelligent, temporarily competent and locally complete).

Method

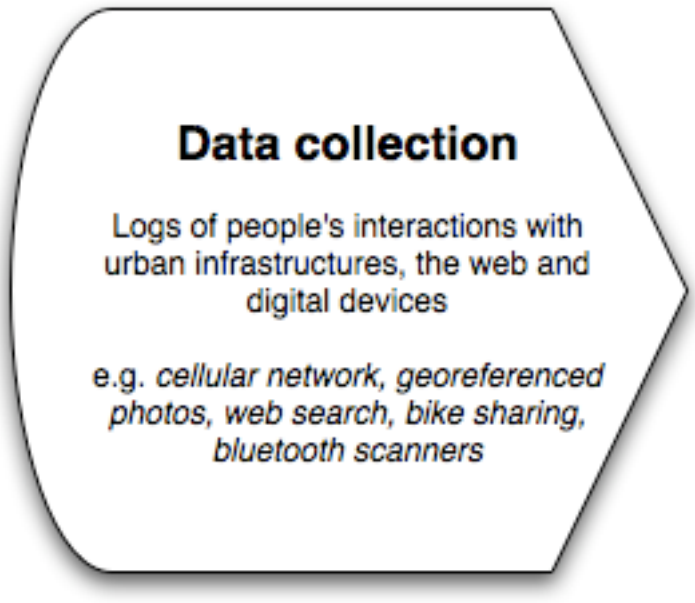


In my research, this is the kind of 3-steps methods I have been using to produce means for sentient and responsive cities. Data collection, observation and indicators.

Applications

If will describe the application of this method through several examples.

Data collection



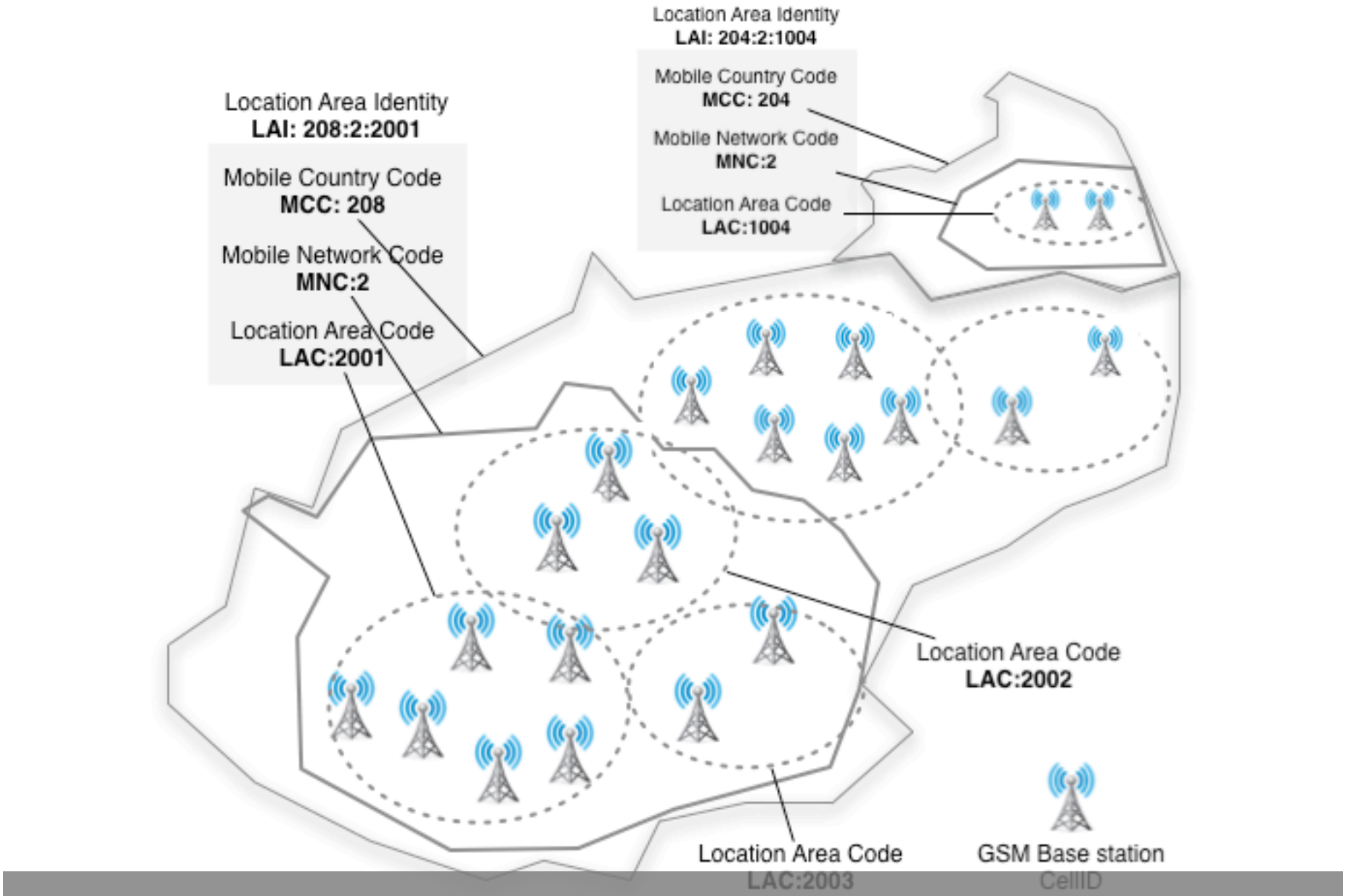
First the collection and use of digital footprints

Air-travel survey

- **Motivation**: Capture air travel experience in situ
- **Method**: GSM network fingerprinting
- **Result**: 97% flight detection rate (74/76)
- **Limitation**: Revision of algorithm according to regulations and social rules

Air-travel survey

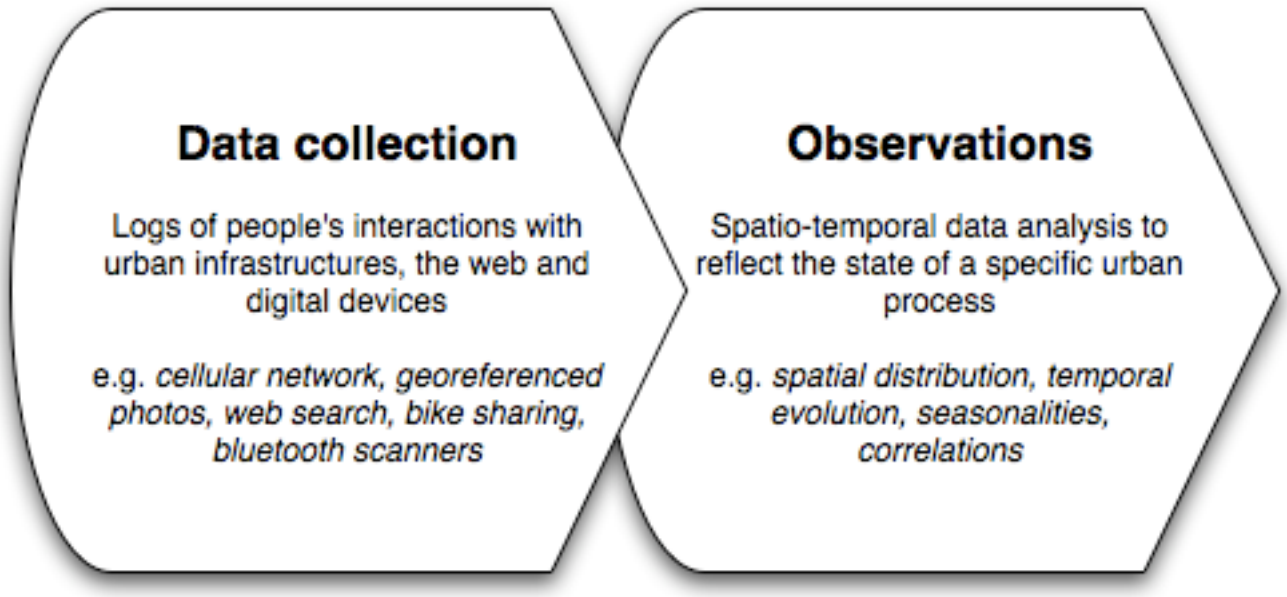
In a project that aimed at capturing air travel experiences when they happened at a world-wide scale, we used the implicit interaction of survey participant's mobile phones with the cellular network. It lead to a 97% flight detection rate (74/76 flights)



LAI fingerprinting

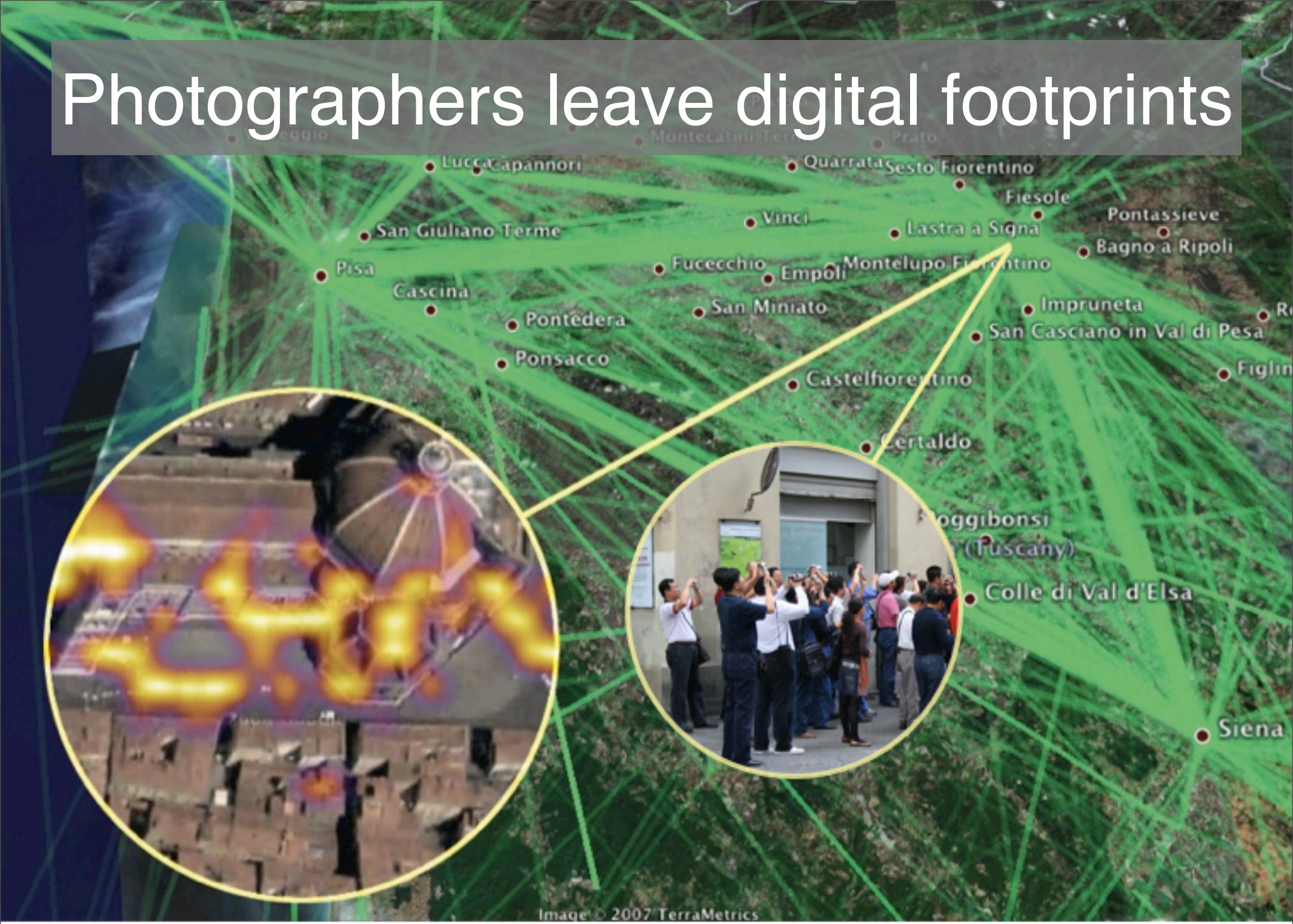
Installed on each participant's mobile phone, our software loged on a regular basis the Location Area Indendity of the area serving the phone. The embedded algorithm monitored the logs to detect motions between. Within special condition, unique to air travel, the algorithm would detect the flight and launch a survey. Privacy of the participant was kept as at no time the embedded software communicated mobility information, unless explicitly consented by the participant.

Observations



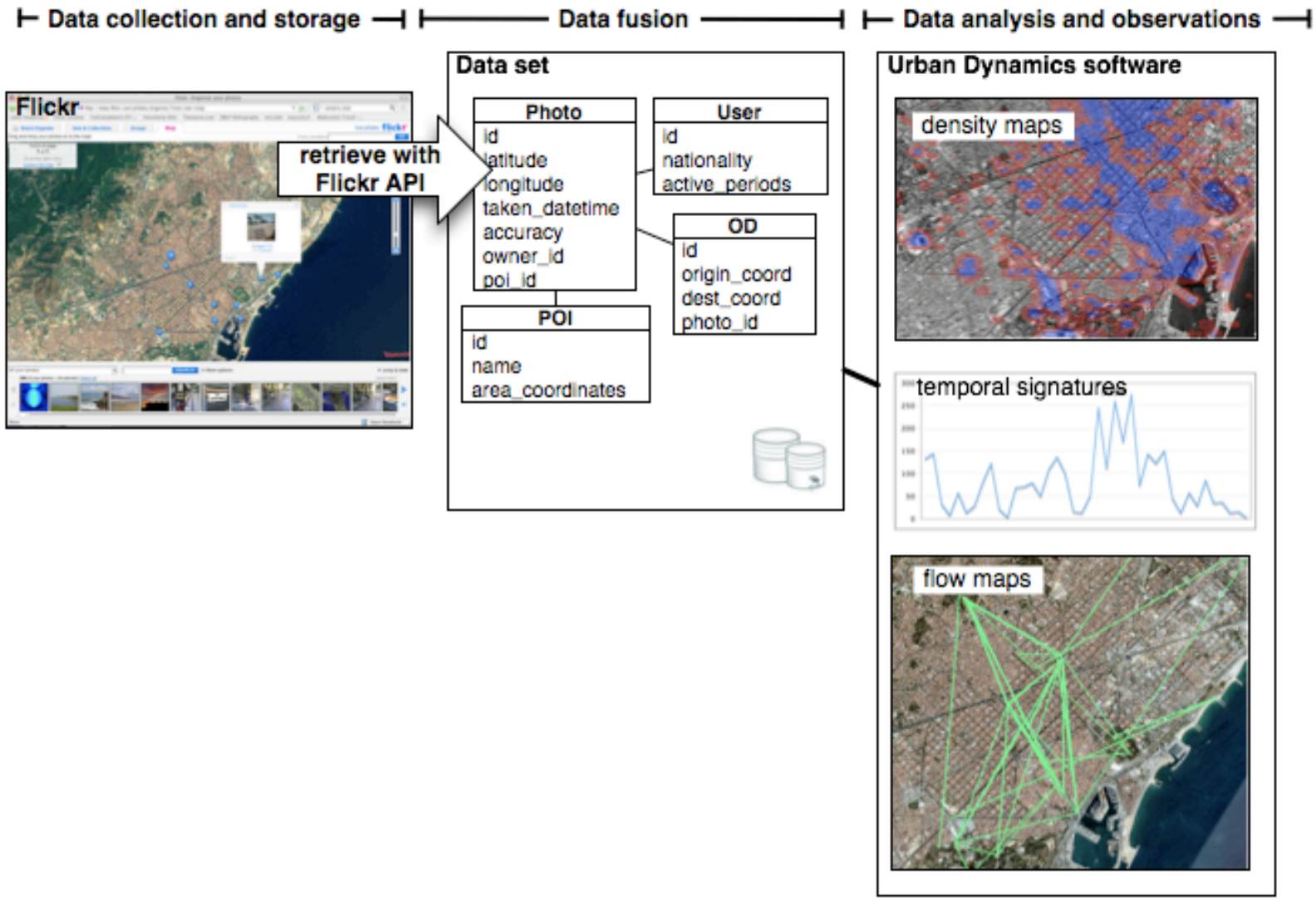
This first example showed how to exploit digital footprints to get mobility information previously tedious to collect. Another opportunity emerges from the observations of larger amount of digital footprints.

Photographers leave digital footprints



In this project, we were contacted by the Province of Florence that is desperate in acquiring information on the presence and flows of tourists who spend the day in the area without leaving any traditional trace of their stay (hotel survey, museum entrance). So we mapped the density and flows of photographers from the content they generate on the web (Flickr)

Method



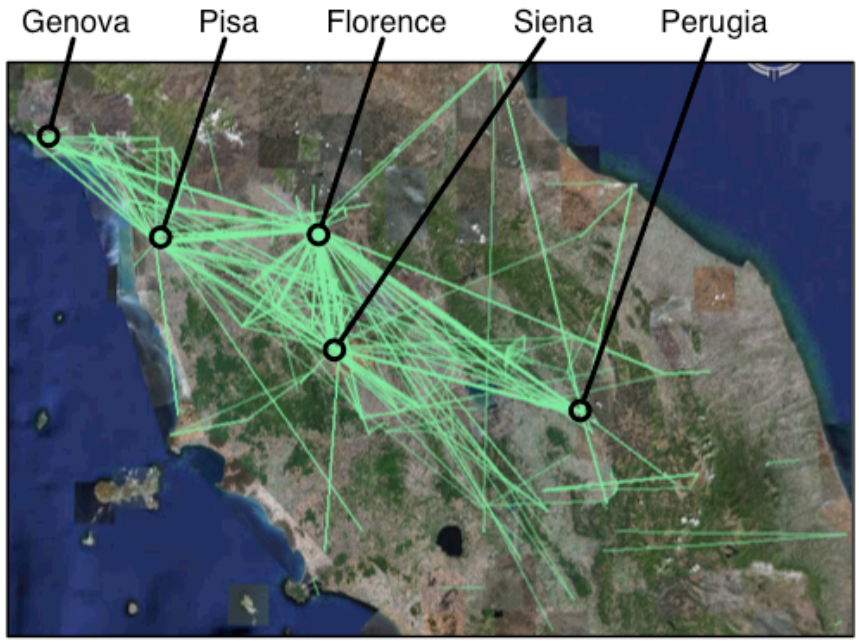
We developed a Urban dynamics software that performs novel data collection and analysis techniques augmented with visualization and mapping tools

Density and scales

Province of Florence (2005-2007)
81,017 georeferenced photos, 4280 photographers

The basic mapping of this data enables to uncovers the density of tourists at very different scales, such as region, citiy and point of interest. While cities are interested in understanding tourist dynamics within their boundaries, we found out that they were equaly interested in understand how the “competitive” cities perform. This type of data allows comparaisons.
Unlike the automatic capture of traces, the manual location disclosure embedded in the act of geotagging of photo provides additional qualities Positioning photo on a map is not simply adding information on its location, it can be considered as an act of communication containing what people estimate as being relevant for themselves and others. In that sense, a specific richness of this dataset arises in the intentional weight people put in disclosing their photos. We show that they have a tendency to select the highlights of their discovery of the city and discarding the downtimes.

Origins and traces

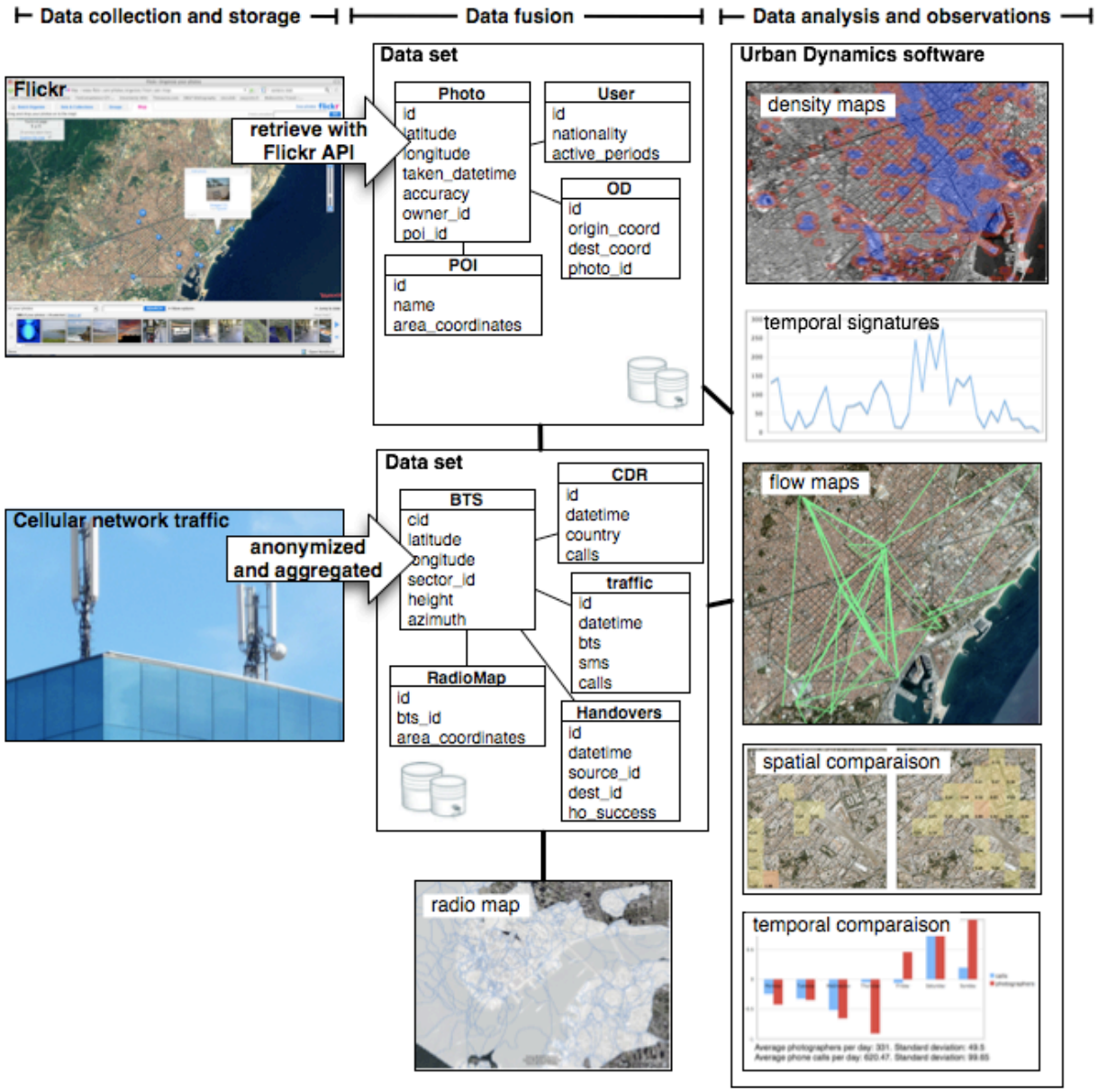


Tourists and locals can be recognized from their practice
60% of users disclose their home country

Flickr lets the opportunity to its users to set their home city or country. We found out that 60% of the users discolsed their home coutry. Taking advantage of this information, we were able to map the traces of american (left) and italian (right) visitors in Tuscany and its proximity. The fact that no American reports a trip to the Island of Elba does not mean they avoid this part of the country, but rather that they do not find relevant to report about it.

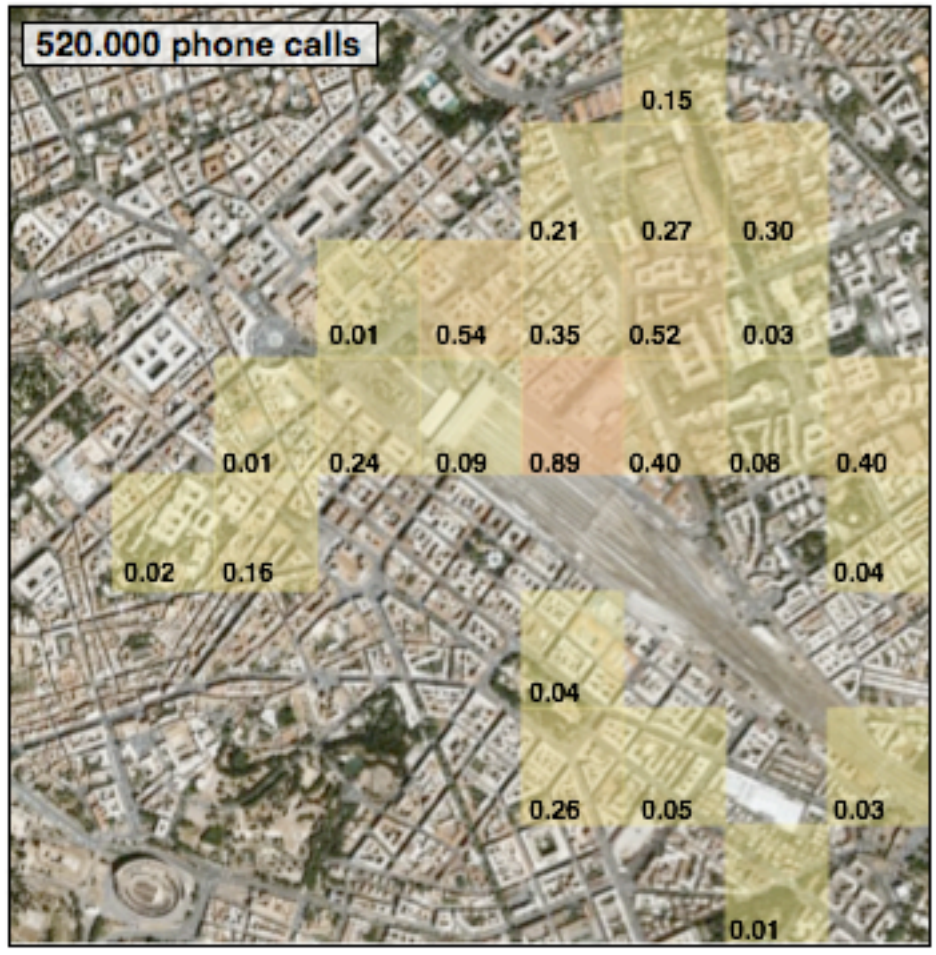
This work in Florence, led us to explore and compare other types of digital footprints

Method



We extended our urban dynamics software to peform the data fusion of multiple data sets, adding visualization and mapping elements to compare the datasets over time and space.

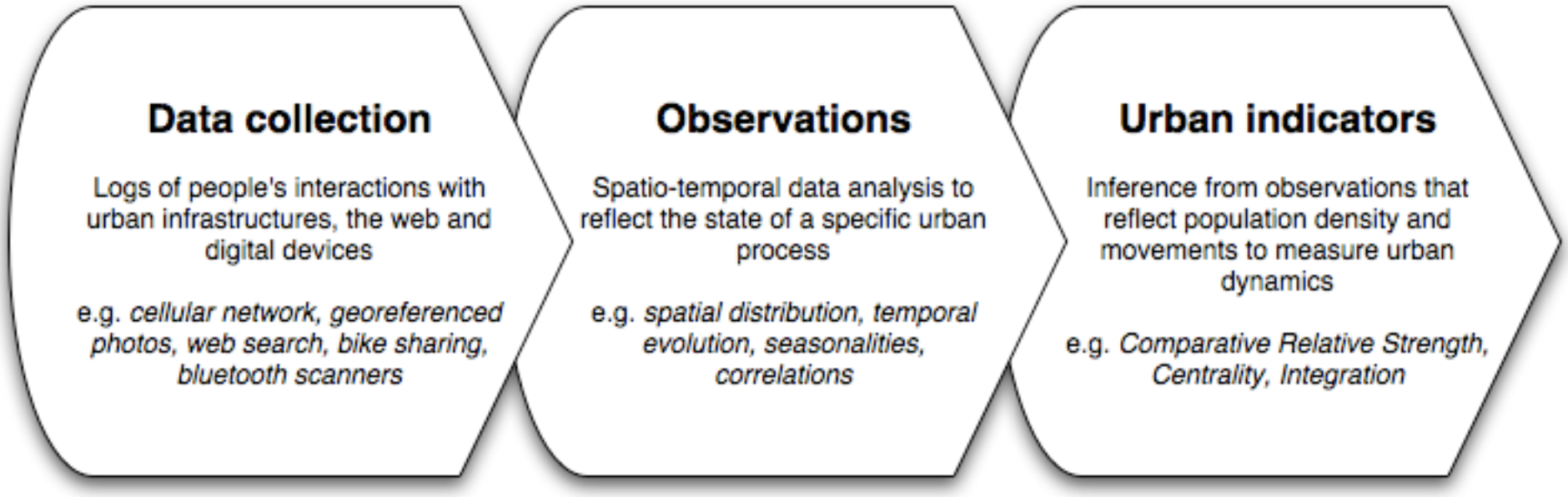
Space and visitors activities



Rome city center (Sept-Nov. 2006)

In Rome, wWe were able to compare the density of tourists from their digital footprints (i.e. georeference photos) and digital shadows (i.e. aggregated cellular network traffic) they generate when visiting Rome. The spatio-temporal analysis revealed that georeference photo were a good proxy to capture the presence of tourists in their sightseeing activites. On the other hand cellular network traffic data were good indicators of space where tourists are on the move or in their accomodations.

Indicators

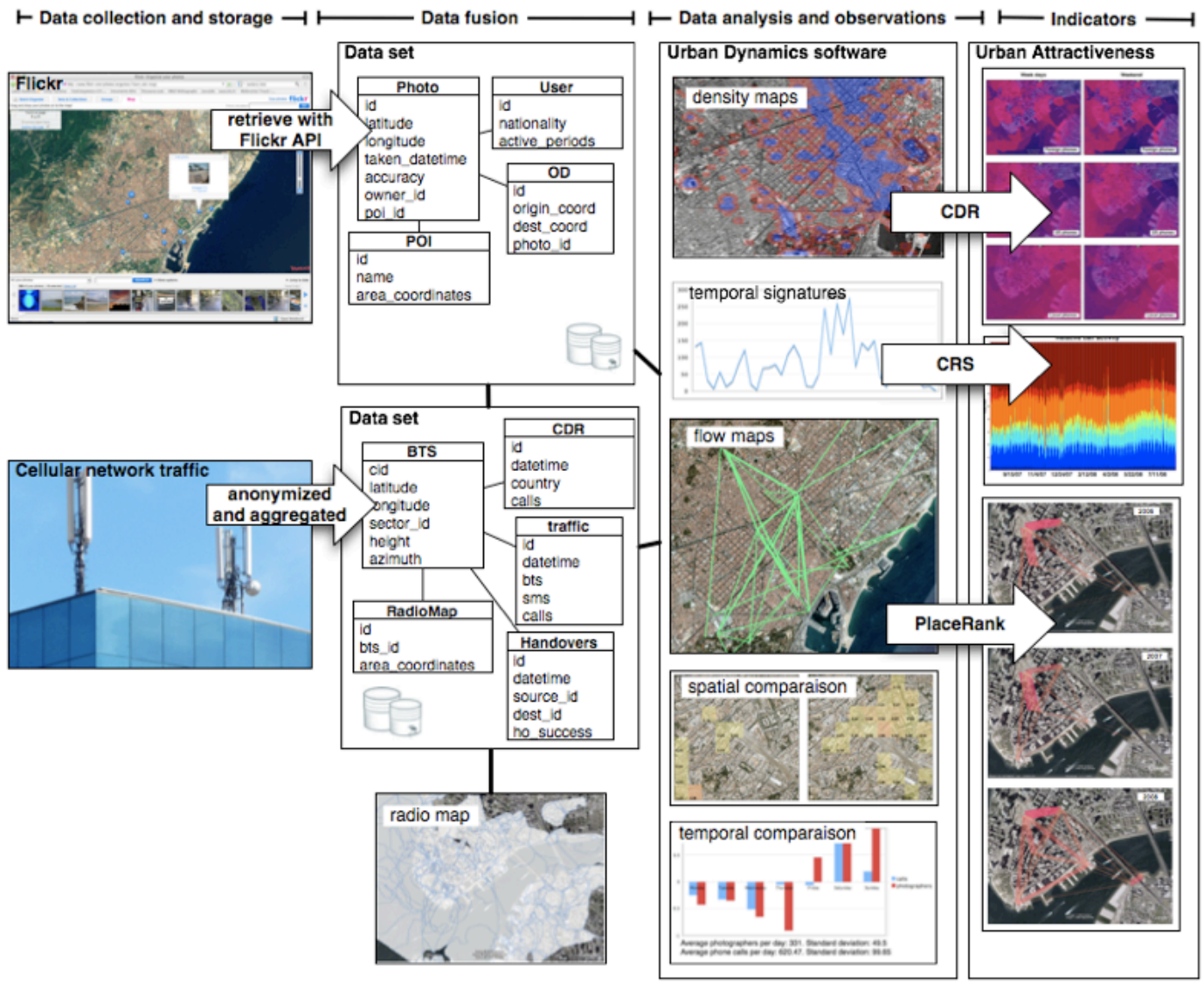


The results in Rome led us to consider the different types of digital footprints are indicators of the attractiveness (are should we say attracTIONess?) of the space.

Attractiveness of the NYC Waterfront

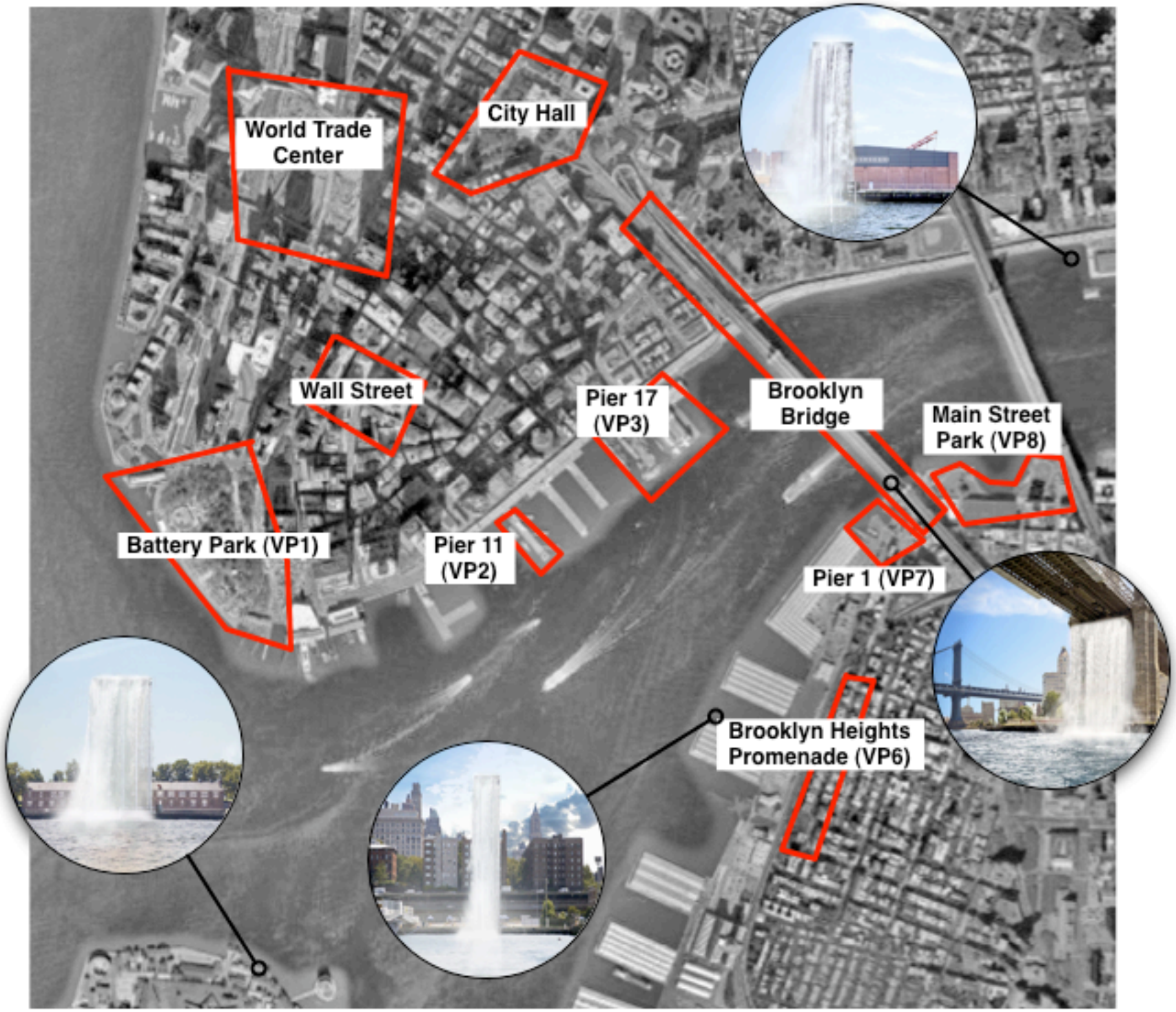
In summer 2008 around the New York City Waterfall public exhibit, we performed further studies on the charactristics of explicit (georeferenced photos) and implicit (aggregated cellular network traffic data) digital footprints to define indicators that measure the evolution of urban attractiveness.

Method

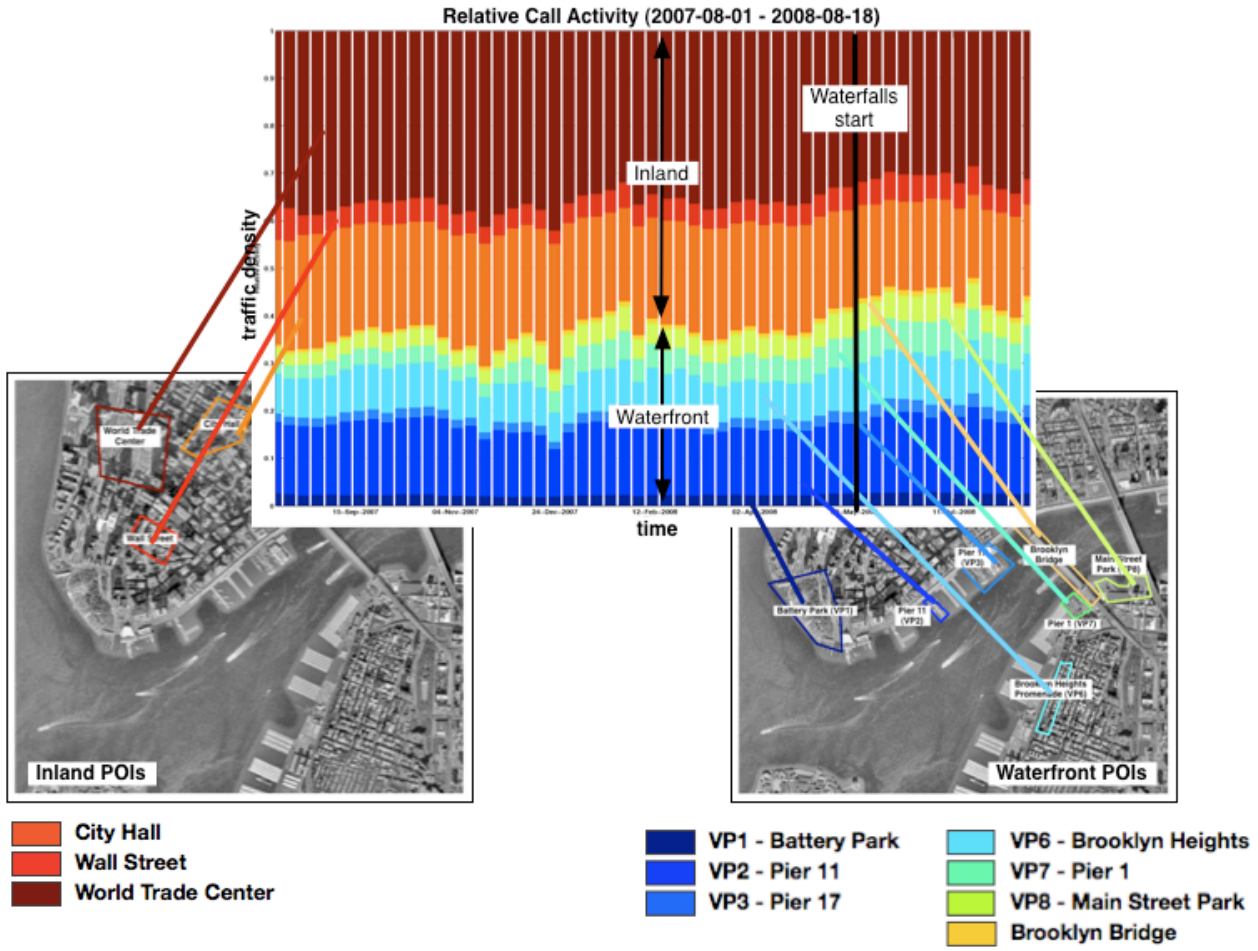


We extended our Urban Dynamics software with algorithms and mapping elements to produce and visualize the indicators

The production of a radio map from a propagation model is an important step to estimate the density of cellular network activity. Here is a video of a very early work that show the evolution throughout a day of the density of the network activity in for each partition of the radio map. It is still extremely hard to produce quality radio maps, particularly in dense urban environments such as Lower Manhattan.


World Trade Center
City Hall
Wall Street
Pier 17 (VP3)
Brooklyn Bridge
Main Street Park (VP8)
Battery Park (VP1)
Pier 11 (VP2)
Pier 1 (VP7)
Brooklyn Heights Promenade (VP6)

Comparative relative strength



The CRS indicator compares the (normalized) activity of one area of interest with respect to the overall activity of the city.

We developed several indicators to compare the evolution of the attractiveness of different areas around the waterfront and of areas of interest in proximity of the exhibit, based on their Comparative Relative Strength (CRS).

PlaceRank

PlaceRank determines the centrality of a location within a set of areas of interest based on the amount of digital footprints generated in each area and the traces that connect them

We assessed the popularity of an area of interest by studying its ties with other areas in the city. The stronger the ties, the more frequently an area is accessible from other places as it becomes part of a popular route. This was measured by applying network analysis techniques to study the connectivity of a network in which the nodes represent areas of interest and the arches represent flows of people between them. Flows were estimated by analyzing concecutive time stamps tagged to Flickr photos in conjunction with the reported location at which a photo was taken. The PlaceRank indicator, inspired by the PageRank indicator developed by Google to order the importance of Web pages, determines the centrality of a location within a set of areas of interest based on the amount of digital footprints generated in each area and the traces that connect them. In particular, if more visitors visit place A than place B, then we can say that the former is more popular than the latter in the network of tourist destinations.

Method limitations

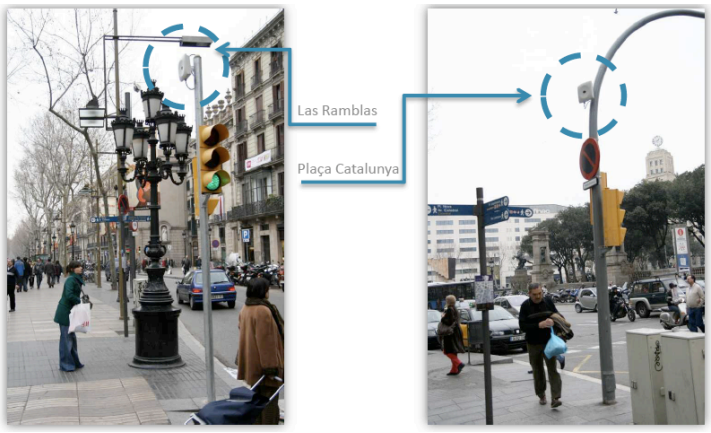
- Reveal phenomena **without explaining them**
- The **extent of their reliability** is still unclear: Lack of callibration with ground truth data (hard to collect and get access).
- **Sense what is cheap to sense**

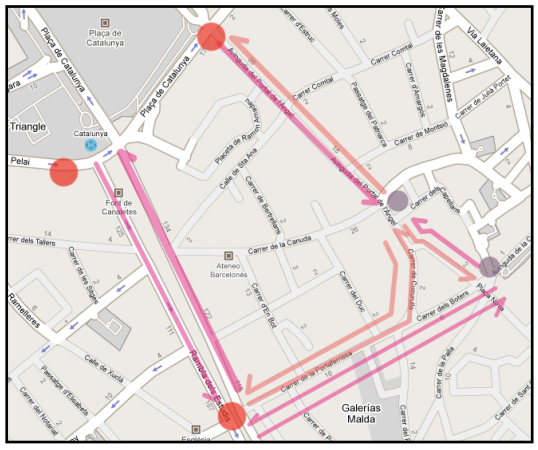
The integration of these results in the official study of the economic impact of the New York Waterfalls public art project shows that our indicators offer useful measures to complement traditional methods. However, the methods I have presensed still contain many limitations. 1) Reveal phenomena without explaining them 2) unclear reliability 3) They can be employed in specific areas (we sense only where digital footprints are massively produced)

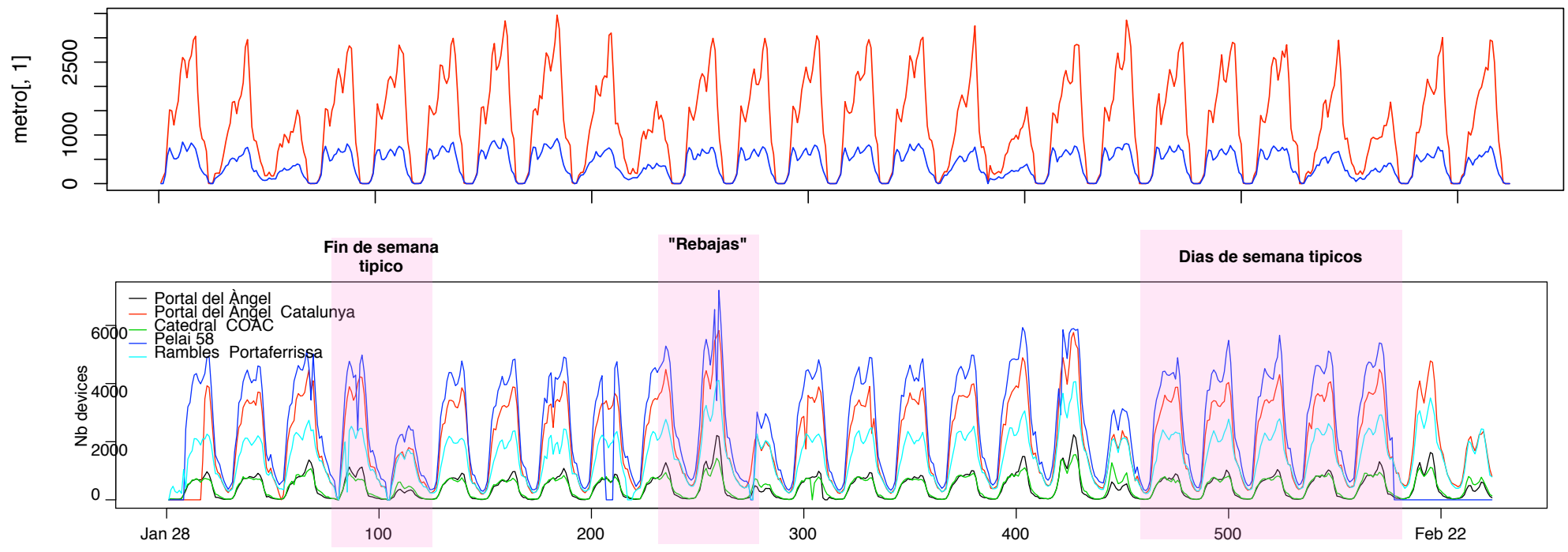
~~Futur~~ Current endeavors

At Lift lab, we took into consideration these limitations, fruits of my PhD thesis, to drill further these urban data layers to create instantiations of possible near future applications.

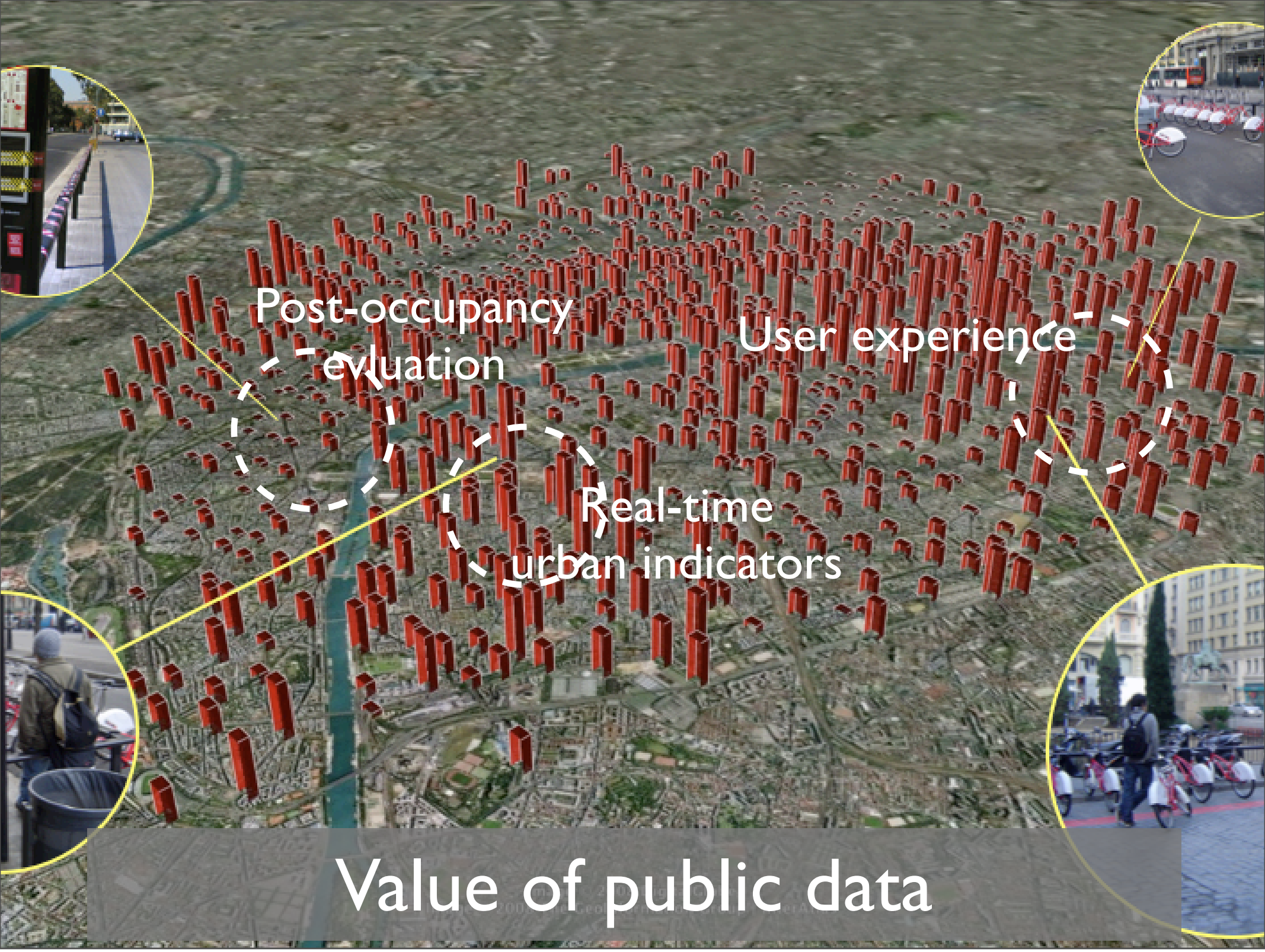
Calibration







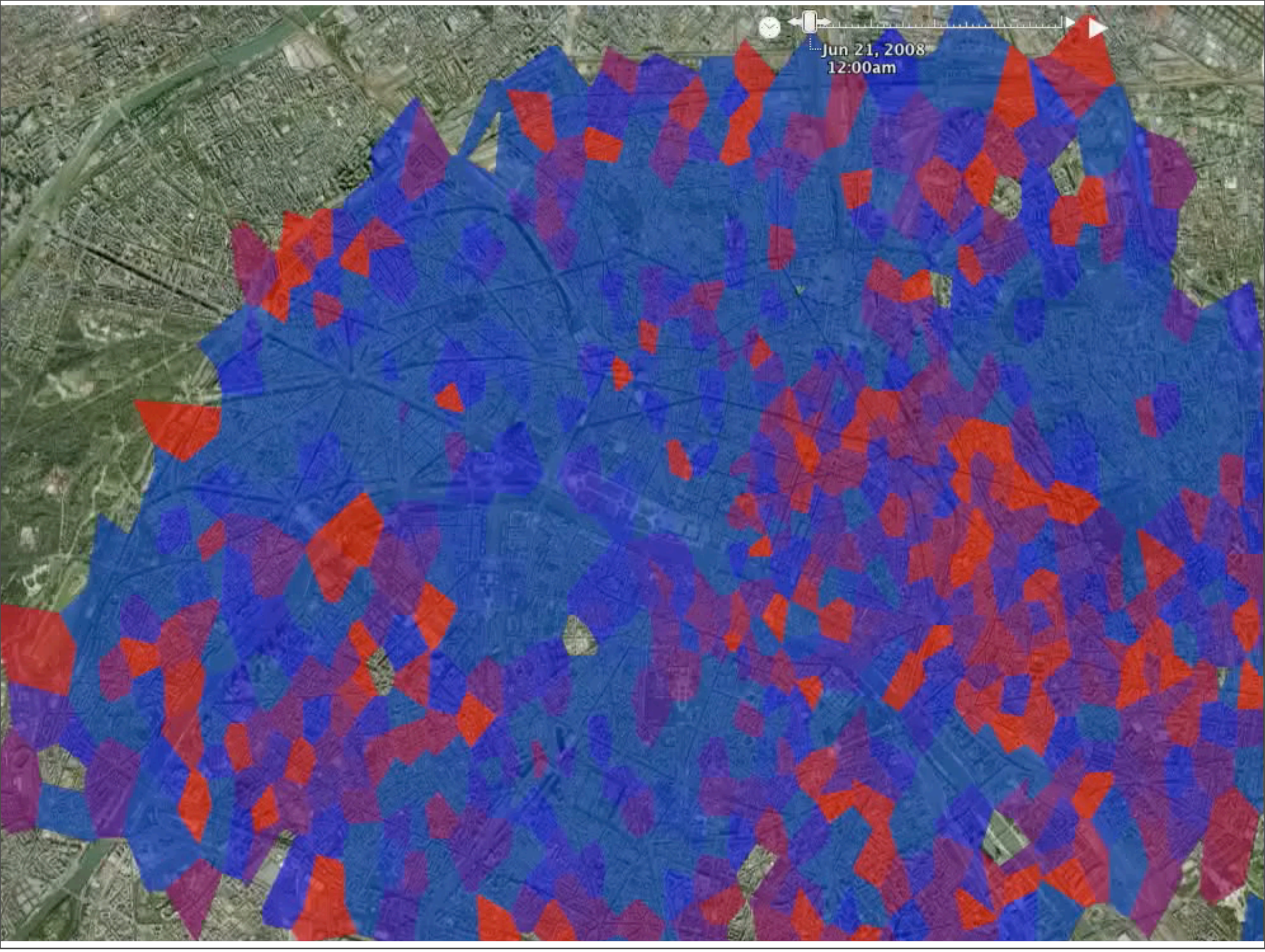
In collaboration with BitCarrier that provides Bluetooth scanning system to monitor traffic, we are able to measure the impact of a sales period on specific parts of a city. This is done with the calibration of bluetooth data with ground truth data. This kind of work aim at finding other utilities to traffic data that are currently collected in contemporary cities.



Value of public data

In the same line of work, we are developping multiple types of indicators from data publicly available such as here Vélib bike station availability data.

The production of a chronotope from this simple data already allow to reveal some elements of the pulse of Paris. http://www.girardin.org/fabien/tracing/velib/



A further chronotop could reveal the hot spot of the “Fête de la Musique” in Paris. Here the replay of the Velib bikes availability during the Fête de la Musique on June 21, 2008 (12:00am to 6am the next day) in Paris. Red areas show stations with a large amount of bikes AND more bikes available than a regular Saturday (comparison with June 14, 2008). The Voronoi Diagram is built from the Delaunay Triangulation. Video: http://www.flickr.com/photos/drremulac/3661877295/

“Citizen science”

- Support citizens in data collection and analysis (participatory sensing)
- Open data access
- Provide mapping and visualization tools to support exploration and argumentation (use the rethoric function of the map)
- Engage in debate and discussion

Another line of investigation aims at envolving citizens into the process to 1) educate on the opportunities and issues around “not so” volunteer generated information and 2) go beyond the sense what is cheap to sense limitation.

Final considerations

- Towards “Observe and improve”
- End of the ephemeral (replay the city)
- Human side of the data (the “intelligent maps” confusion)

Some final consideration on the work and ideas that I have presented in this talk. With an utilitarian perspective, we are trying to move the management of cities from “predict and accommodate” to “observe and improve”. The data collected, the maps and chronotopes developped show some sort of “end of the ephemeral. We are able to replay the city, as oligopticons can rather than a panopticon would. Beyond the utilitarian perspective, there is a big assumption in seeing the world as consisting of bits of data that can be processed into information that then will naturally yield some value to people. In consequence, let us not confuse the development of novel maps from previously uncollectable and inaccessible data with the
possibility to produce “intelligent maps”.

Thank you and thanks to Pablo de Soto and the rest of the organizors for their invitation.

More on http://www.liftlab.com/think/fabien