Network data as material to shape urban strategies

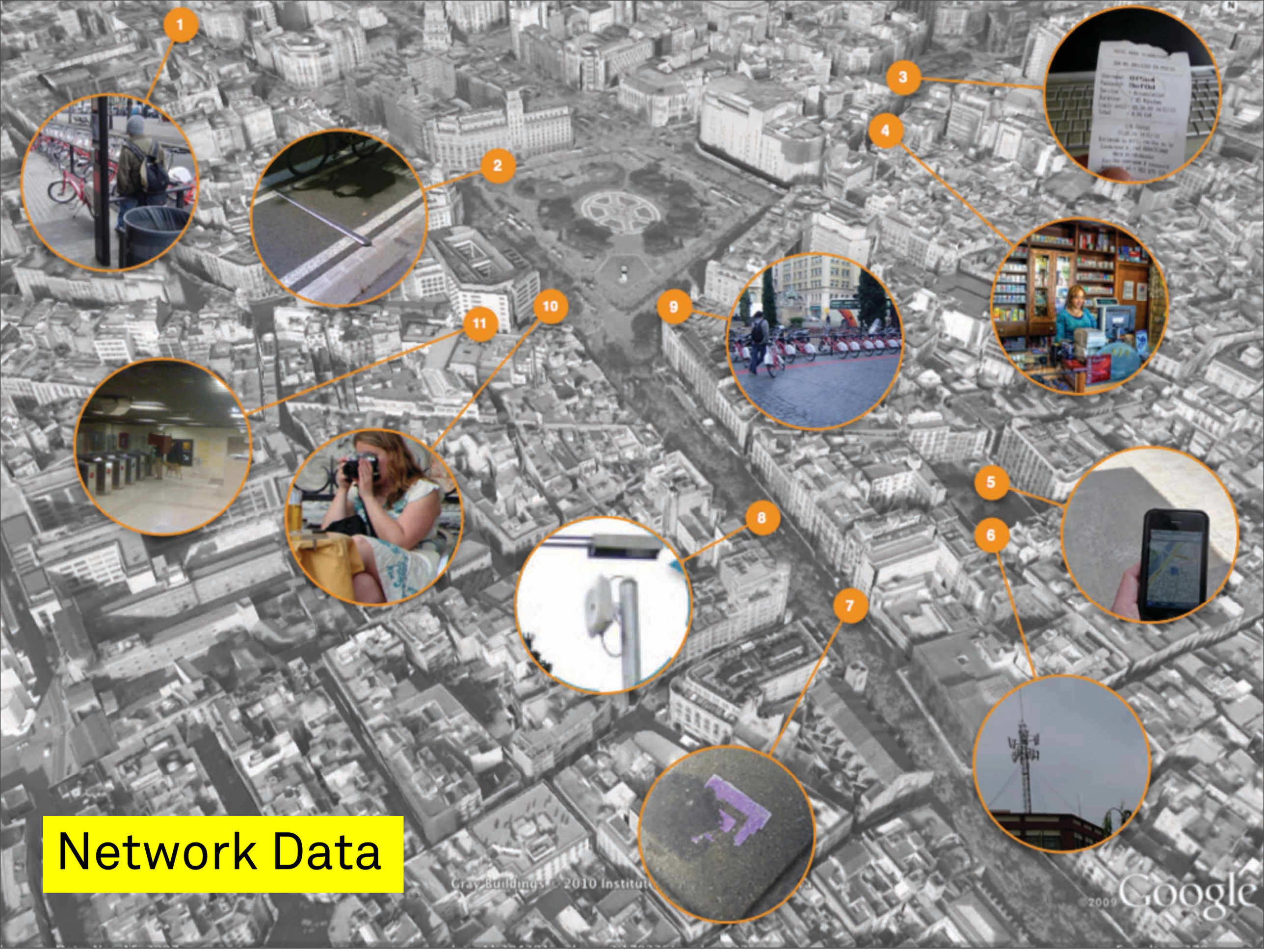
Fabien Girardin. Lift France 10. Marseilles, July 6, 2010.

This talk grounds on the research body I developed at part of my PhD thesis at Universitat Pompeu Fabra and MIT. My investigation encompasses both the physical and the digital forms of the cityscape and their relationships with mobility and the use of space. Now at Lift Lab http://www.liftlab.com/, a research agency I co-founded last year, we are engaged with city governments, urban service providers and space managers who request “operational” outcomes at best affecting their strategies at worst refining the understanding on mobility and occupancy levels. Particularly, we have been developing and applying techniques to extract value from network data with the prerequisite to protect individuals right to privacy.

Let me bring you back in time to provide some context. Last year at Lift France, Timo Arnall presented his extensive work on near-field communications (see Making Things Visible http://www.liftconference.com/fr/timo-arnall-quotmaking-things-visiblequot.

Wireless in the World

Timo showed us how wireless technologies are now embedded into our objects, our streets and our experience of the city. See Wireless in the World 2: http://www.nearfield.org/2010/06/new-film-wireless-in-the-world-2. This evolution and its implications raised my interest. Our constant interactions with network infrastructures generate data that aliment the integral function of services we use in our daily lives.



These network data are the result of interaction with the wireless, digital and software infrastructures that are now part of the cityscape. Some we are aware of, some were are not. I encourage you to pay attention to the situations you actually generate network data, examine what they are and where they are stored. (see Adam Greenfield and Nurri Kim's Systems/Layers walkshops http://speedbird.wordpress.com/2010/05/10/how-to-bring-a-systemslayers-walkshop-to-your-town/)

Here I condensed the multiple frictions with sources of network data in my home town of Barcelona. Within a 40min time span, 1) I checked out a bike from the Bicing network using my member card, 2) I rolled over a loop detector 3) I accessed a coffee shop WiFi network 4) a cashier swiped my fidelity card 5) I checked-in Foursquare and access Google Maps on my mobile phone 6) As I am moving, two mobile network antennas produced a hand-over of my communication 7) I acquired a product with an RFID attached to it 8) I pass nearby Bluetooth scanners deployed to measure the traffic at La Rambla 9) I checked back in my bike 10) I appeared in a photo taken by a tourist at Plaza Catalunya, later uploaded to Picassa and finally 11) I use a T-10 ticket to access the subway system.

All these data are basic element of the functions of urban systems, services and infrastructures that archive them.

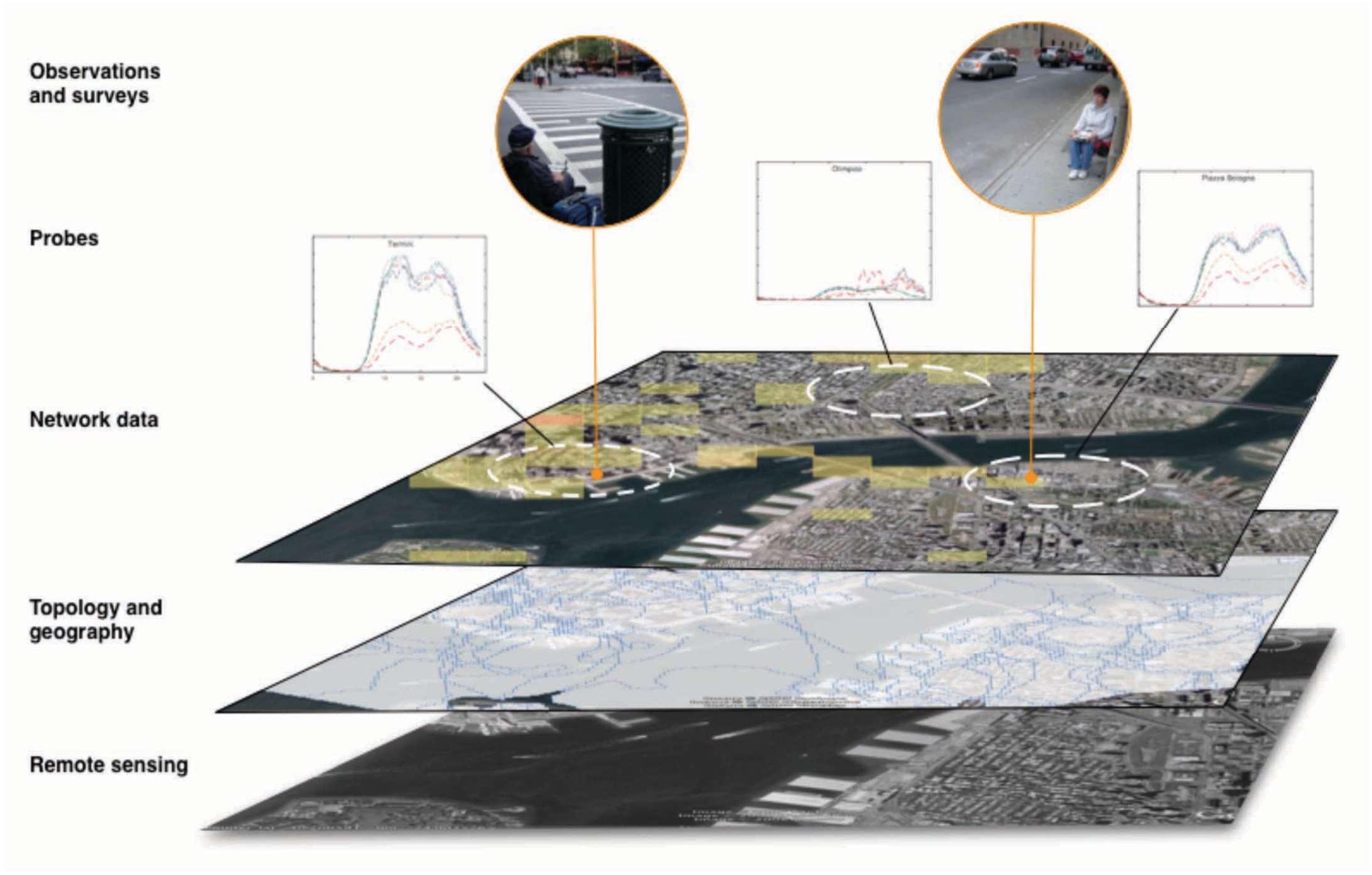
The exploitation of network data

People services
Business & Commerce
Mobility
Communication
Energy
Entertainment

In many ways these data are becoming the “lifeblood of today’s economy” (see Daniel Kaplan’s Digital Privacy Revisited http://www.slideshare.net/slidesharefing/digital-privacy-revisited).

We are, of course, not the unique institution interested in trying to extract value from this massive amount of data in the context of the city. Some leading players of the ICT and engineering world such as IBM, CISCO and Arup and leading major initiatives (Smarter Cities, Connected Cities, Networked Cities, ...) strongly based on data and their analysis to innovate in the domains (often described as “systems”) of public services, business, mobility, communication, energy and entertainment). The projects I present in this talk certainly lay at the crossroads of these domains.

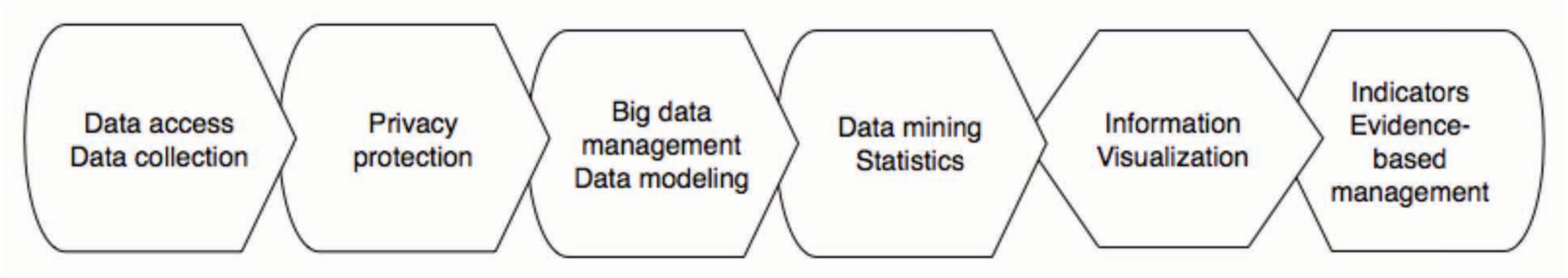
Materials to shape urban strategies



Practically, there are multiple sources of material to shape strategies in these domains, from the information on the physical aspects of the urban environment with remote sensing techniques augmented with topological (networks, infrastructures,services) and geographical information. Network data provide a more dynamic evidences of urban activities with the ability to probe network activities in specific areas for specific periods of time. My talk will focus on this specific layer.

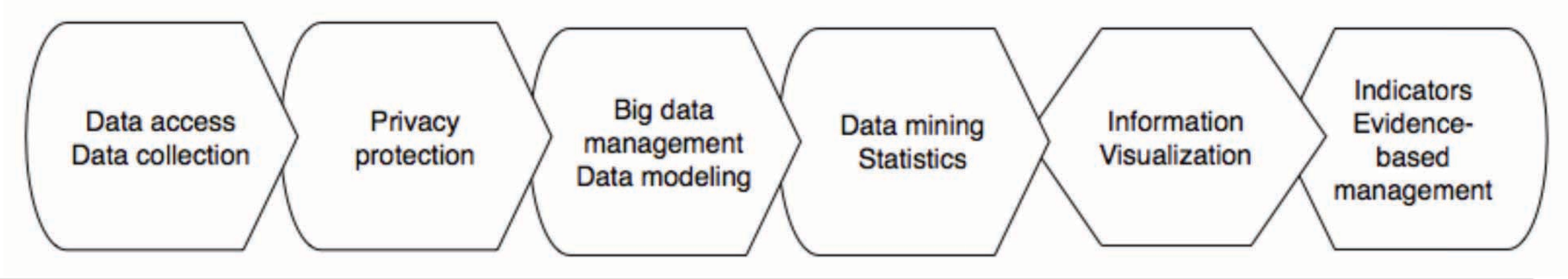
Finally, qualitative observations on the field and surveys form also part of the material to understand an urban environment and shape strategies.

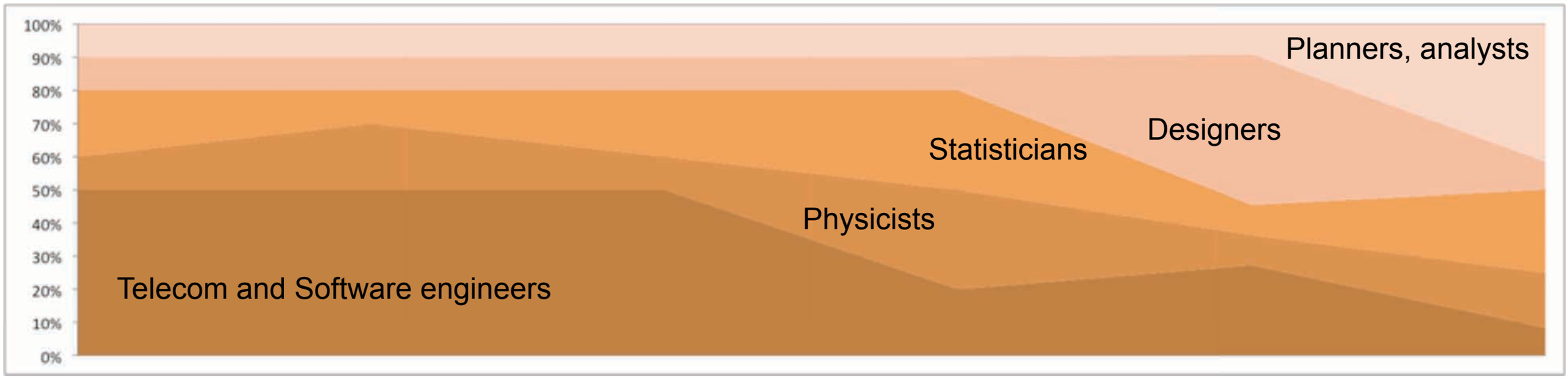
The materialization process



The process to generate material from network data takes a few steps and iterations.

The data scientist profile





A major challenge in this process consists in applying the necessary skills for each phase, starting with Telecom and Software engineers techniques to collect and access network data and apply the necessary algorithms and processes to protect individuals' right to privacy (e.g. anonymization, encryption, obfuscation, aggregation, deletion). Besides data modeling capabilities of software engineers comes the long experience of physicist to extract information from the noise naturally embedded in massive amount of data. Skills in statistics is compulsory to produce and validate information. The particular kind of information that designer know to visualize and communicate to transdisciplinary audiences of planners and analysts (and by extensions citizens, decision-makers, lawyers, lobbyists...).

The art of visualization

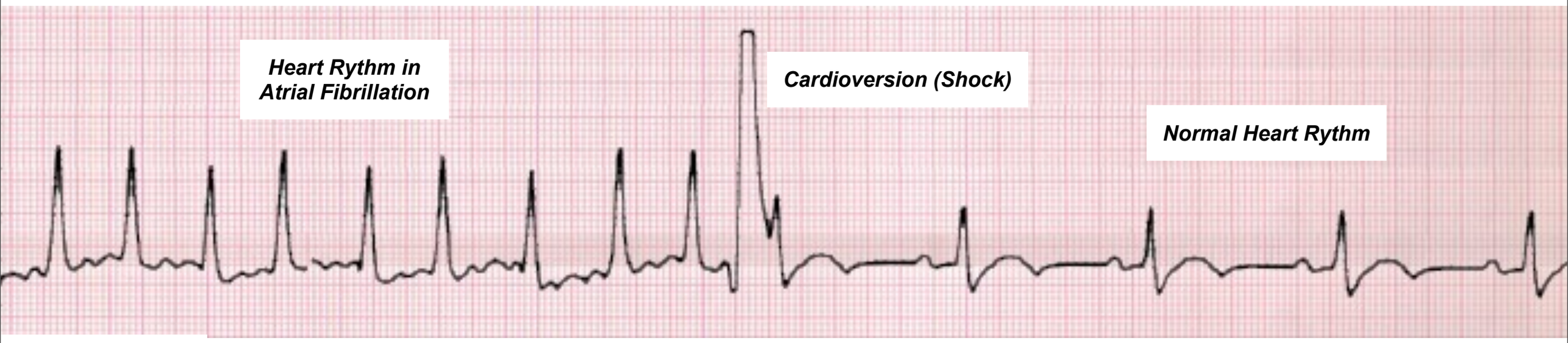
This skills distribution plot by phases is a bit crude, I agreed. To exemplify it a bit more, let me highlight 3 types of “arts” I believe are crucial in the application of data science in the realm of urban strategies.

First, the art of visualization helps engage dialogs and open debates on the quality of the data and their potential value. Here is example my former research lab developed. The MIT SENSEable City Lab project WikiCity http://senseable.mit.edu/wikicity used large screens deployed in the streets of Rome to display in real-time the presence of crowds from their activities on the Telecom Italia mobile phone network overlapped with information on the distribution of the public transport services.

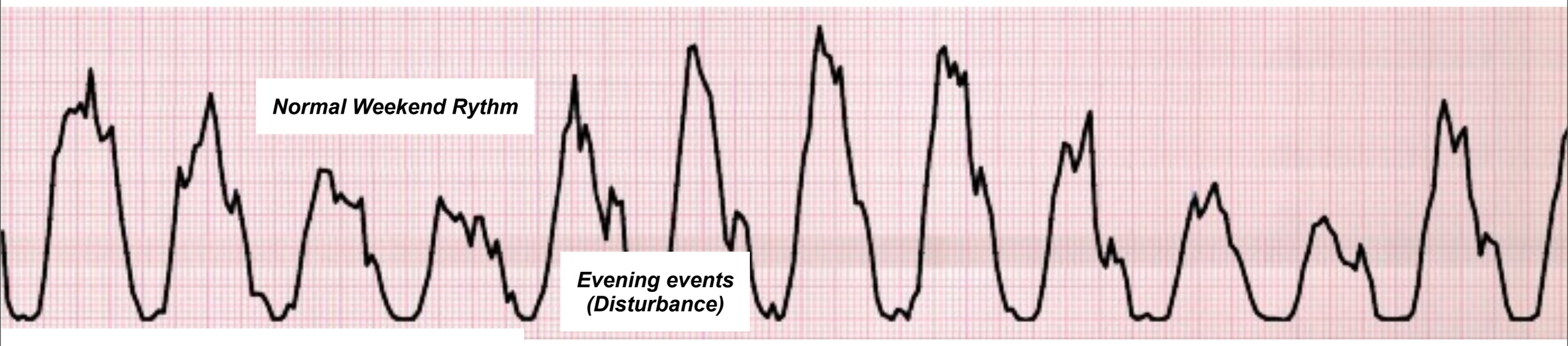
This type of urban demo approach with attractive information visualization is essential to surface a problems, convey a message to a large audience and collect feedback.

As valuable information visualization is to explore and communicate, it is limited when it comes to extract the exact value of raw data.

The art of taking a pulse



Heart rythm

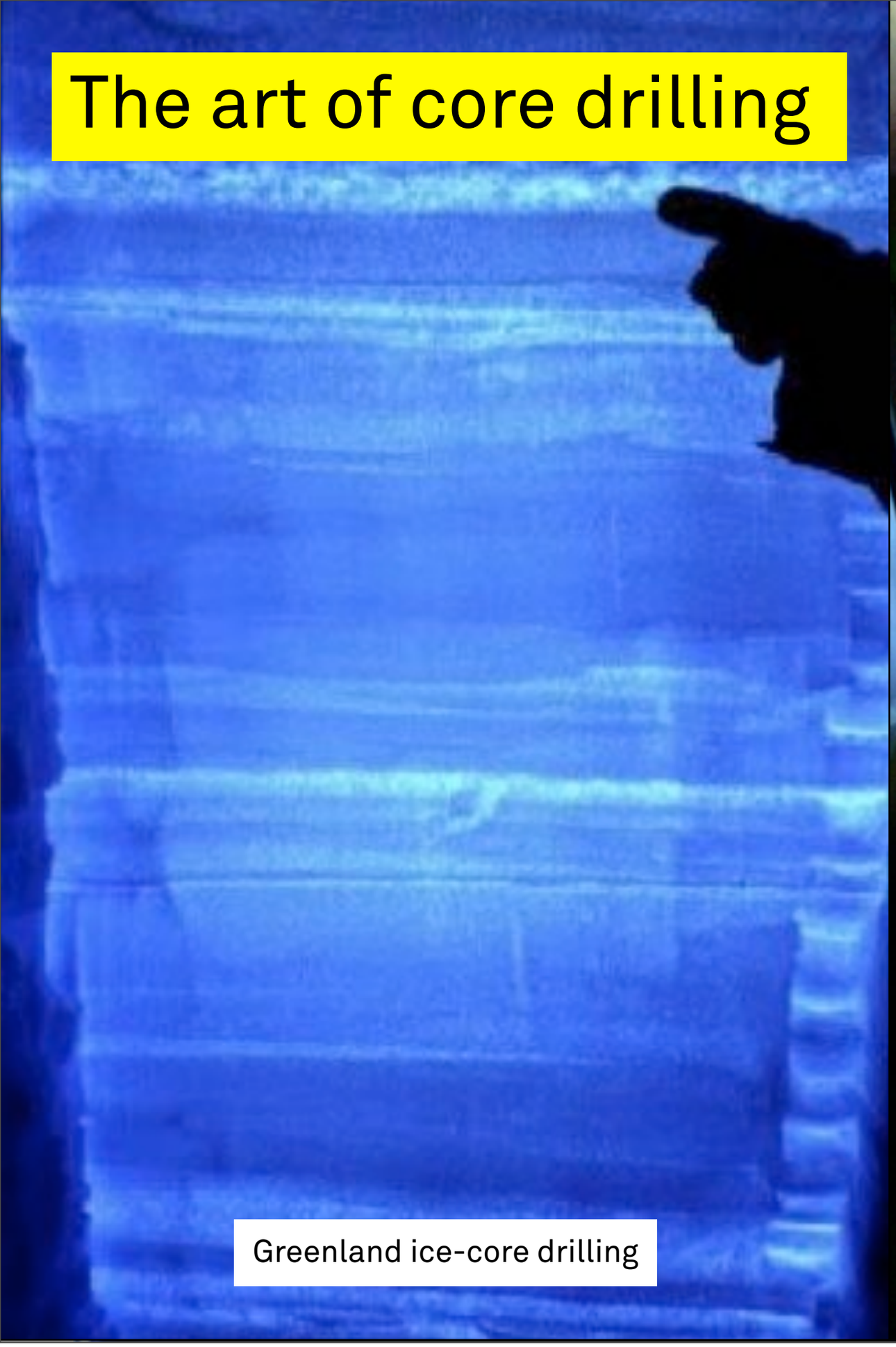


Mobile phone network antenna rythm

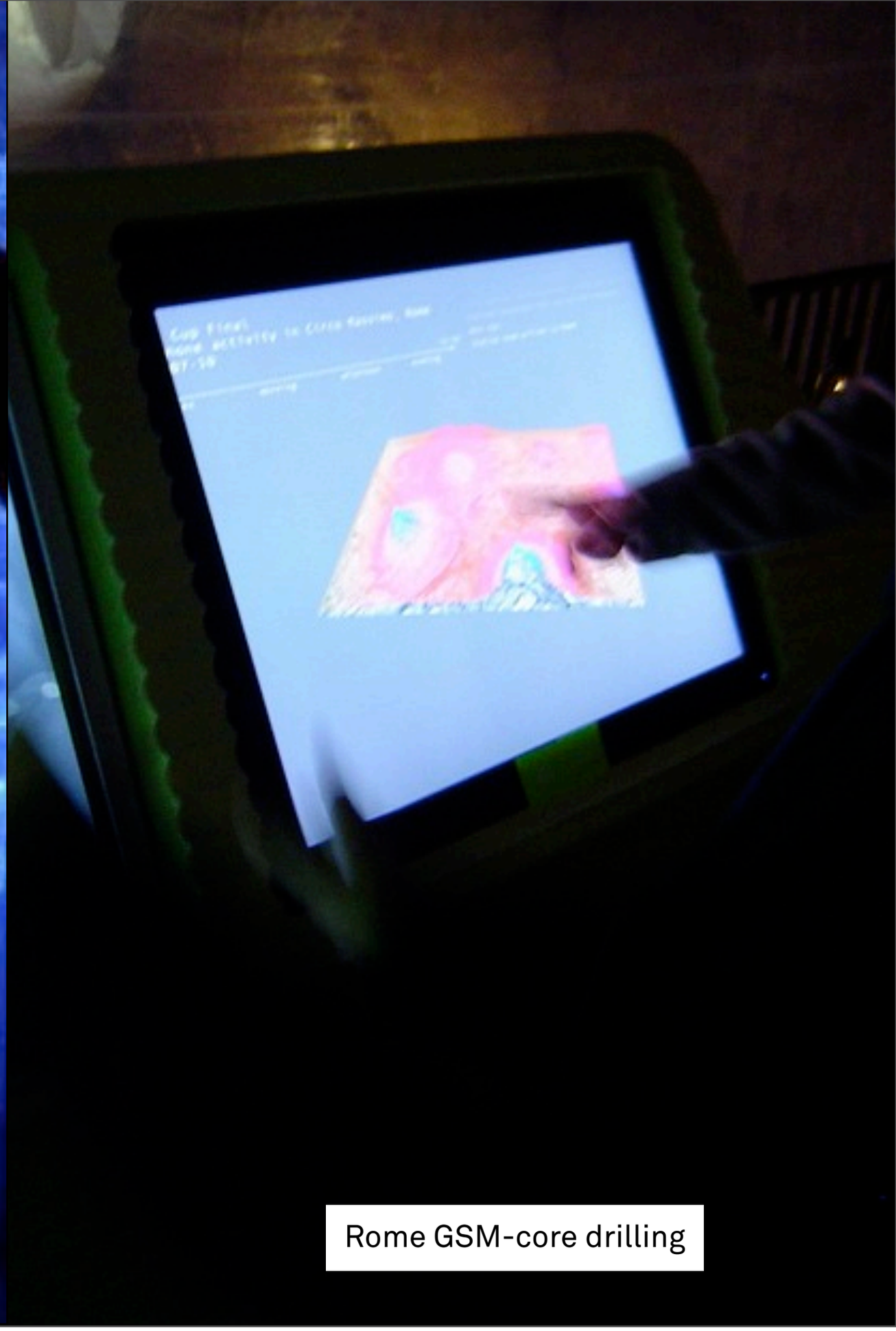
The art of taking a pulse, helps understand patterns and rhythm of the space from its network activity. Comparable to a medical doctor who checks the rhythm of a heart to measure symptoms and external events (here on top with a Cardioversion to convert an abnormally fast heart rate to a normal rhythm), network data infrastructures have their normal rhythm and disturbances from external event (here at the bottom the pulse over 12 days of a Mobile phone network antenna reveals disturbances in on Monday and Tuesday evenings after the first weekend).

The analysis of a pulse often focuses on the detection of patterns and anomalies from real-time data feeds.

The art of core drilling



Greenland ice-core drilling

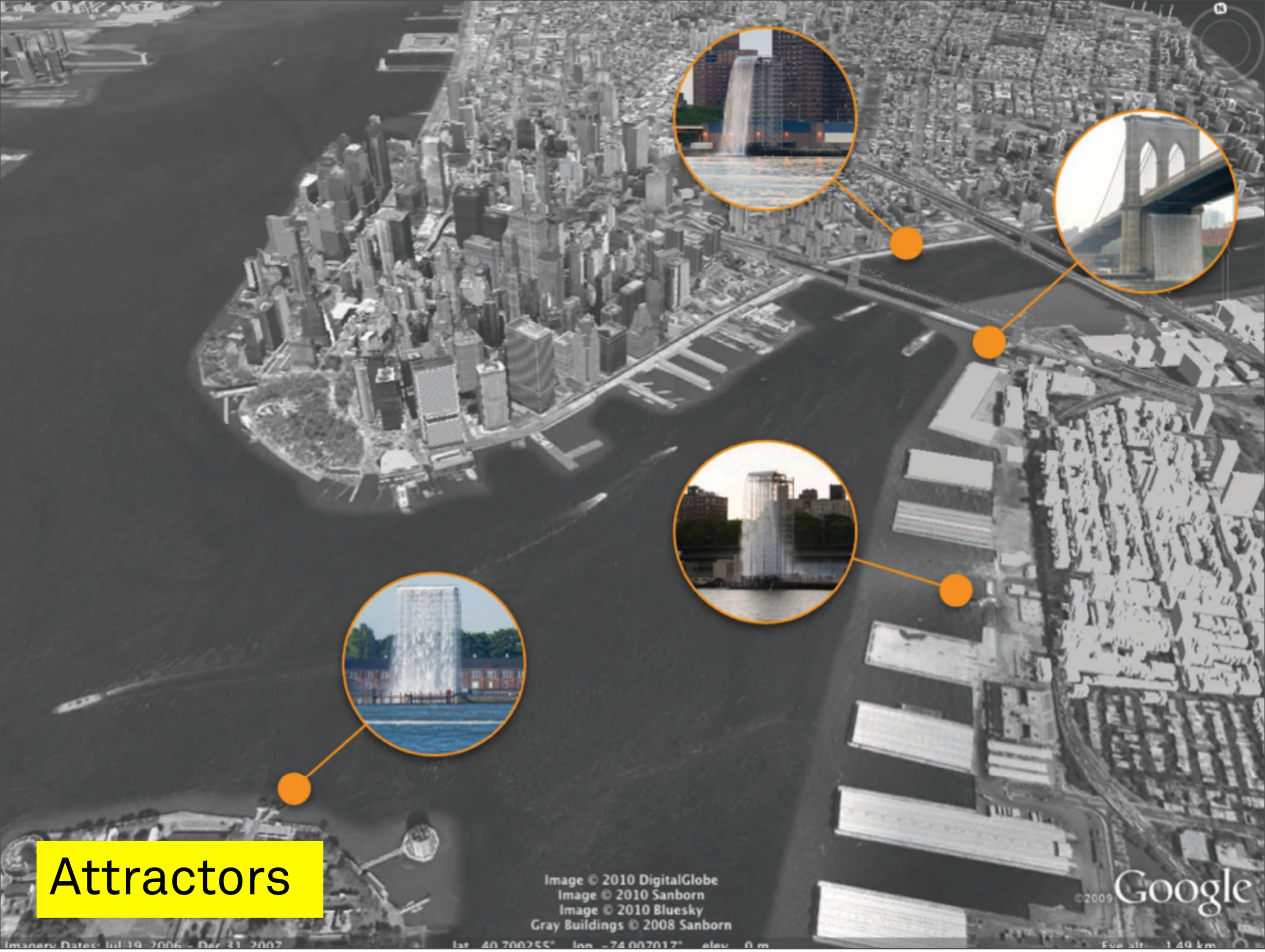


Rome GSM-core drilling

Beyond real-time data, there is an enormous value in understanding the evolution of a space from the accumulation of layers of network data. Comparable to Environmental Scientists who use the art of ice-core drilling in Greenland to analyze the historical conditions of our planet in order to predict its evolution, it is possible to probe the evolution of the network activity for a certain period of time.

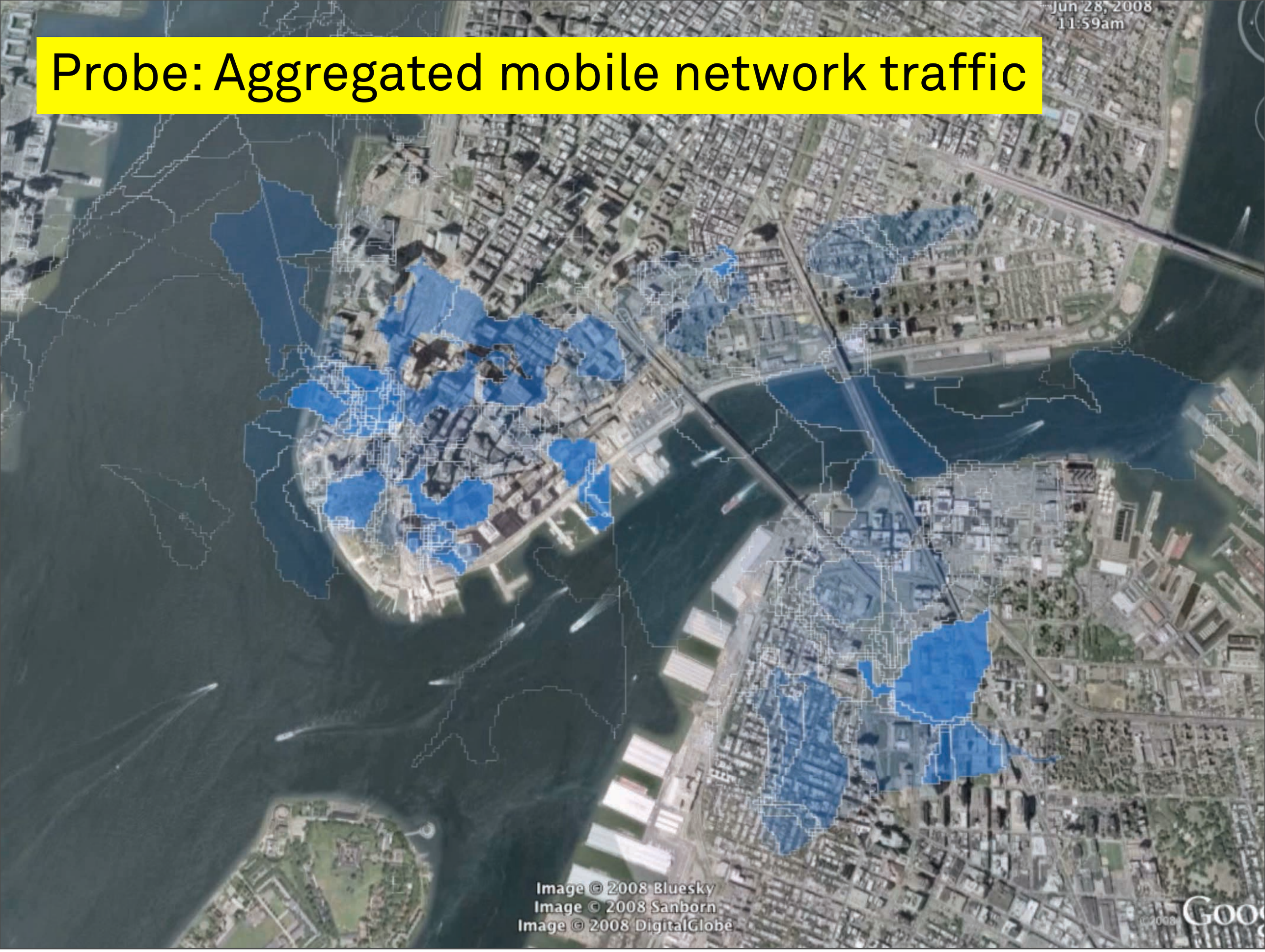
Now let's shift to the operational level. I will introduce you a couple of projects that I believe exemplify the exploitation of network data as material tho shape and evaluate urban strategies.

For this first project, let me bring you bak in time once again. In 2008 the New York was completing several projects to remodel and refurbish the East River waterfront in the Lower Manhattan and West Brooklyn areas. This goal was to make these spaces more attractive to citizens and visitors.



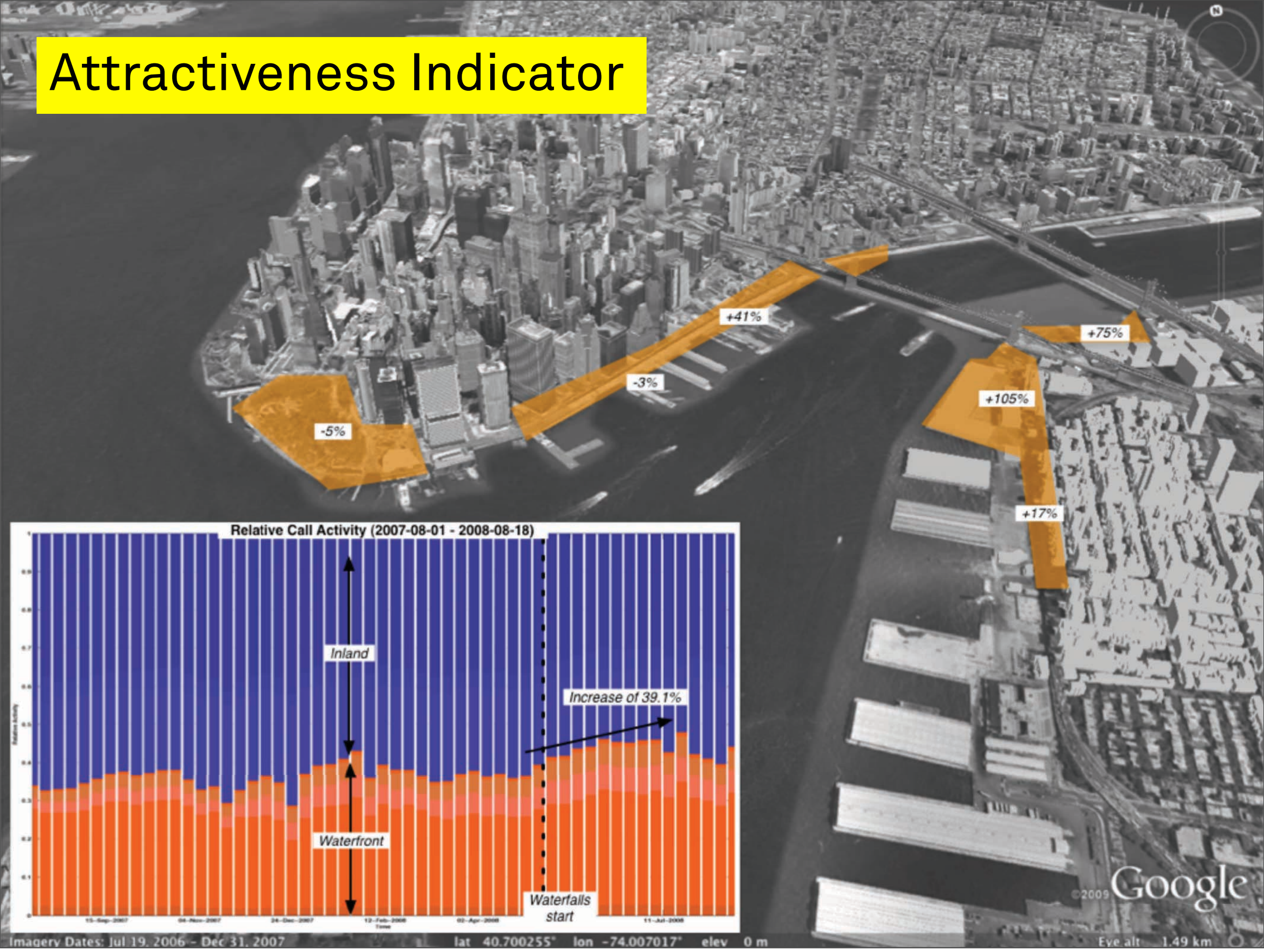
In Summer 2008, as part of the strategy to invite citizen and visitors in discovering and re-appropriating the waterfront, a public art exhibit made of 4 man-made Waterfalls. The New York City Waterfalls that costed around $15mio of mainly private investments necessitated measures and metrics to evaluate its impact.

Probe: Aggregated mobile network traffic



We (MIT SENSEable City Lab) used a couple of probes to measure the evolution of the network and digital activity in the area of the Waterfalls. First we measure the evolution of the mobile network traffic over 1 year starting from the Summer prior to the Waterfalls.

Attractiveness Indicator



In collaboration with data scientist Andrea Vaccari http://andreavaccari.com/, we developed an attractiveness indicator based on the comparison of mobile network activity at the waterfront (the multiple vantage points of the NYC Waterfalls) with areas of the city that have no relations with the Waterfalls. This approach inspired with the technical analysis of financial market (see Relative Strength Index http://en.wikipedia.org/wiki/Relative_Strength_Index). Since then other have inspired from our work (see Skyhook Wireless SportRank http://www.skyhookwireless.com/spotrank/)

When the exhibited opened our indicator revealed an relative increase of the network activity at the Waterfront in comparison to inland areas. This evolution from one summer to another testifies of an impact of the strategy on the presence of people from their activity on the network.

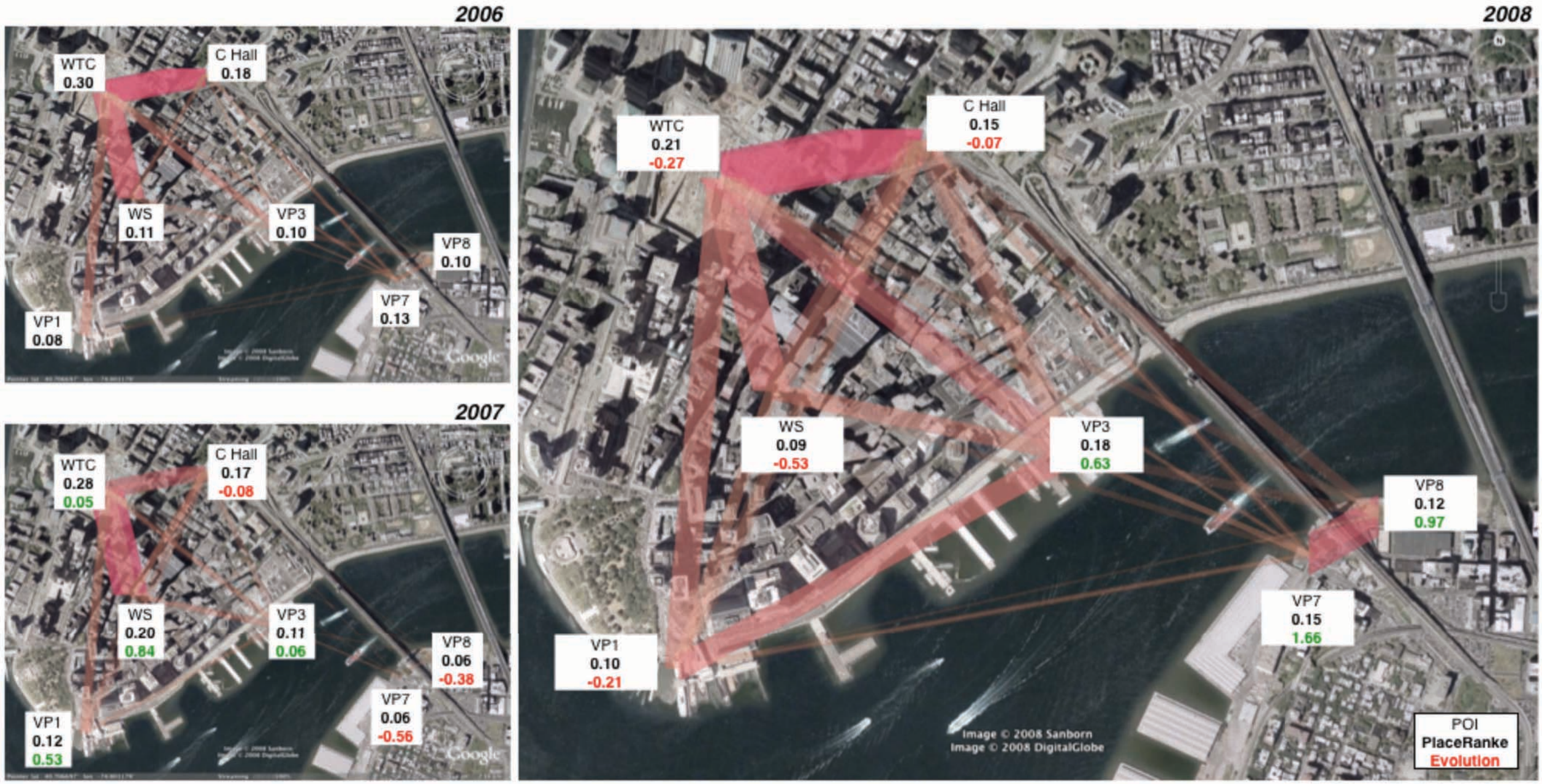
For more detailed description about the investigation, I invite you to read:

Girardin, F., Vaccari, A., Gerber, A., Biderman, A., and Ratti, C. (2009). Quantifying urban attractiveness from the distribution and density of digital footprints. International Journal of Spatial Data Infrastructure Research, 4:175-200.
http://ijsdir.jrc.ec.europa.eu/index.php/ijsdir/article/view/147/152

Probe: Flickr photos

We also looked at the evolution of the photographic activity over three years in Lower Manhattan and West Brooklyn. The dataset, of publicly available elements, was collected from the Flickr API http://www.flickr.com/services/api/.

Centrality Indicator



Based on the location and timestamp attached to public photos, we were able to extract the main flows of photographers in the area. With the application of a PlaceRank algorithm developed in collaboration with Andrea Vaccari (inspired from Google PangeRank http://en.wikipedia.org/wiki/PageRank) the evolution of the flow becomes measurable. In comparison with the previous years, the waterfront becomes more central to the flows of photographers in Summer 2008. Once again testifying of an impact of the strategy on the presence of people from their activity online.

For more detailed description of the investigation, I invite you to read:

Girardin, F., Vaccari, A., Gerber, A., Biderman, A., and Ratti, C. (2009). Quantifying urban attractiveness from the distribution and density of digital footprints. International Journal of Spatial Data Infrastructure Research, 4:175-200.
http://ijsdir.jrc.ec.europa.eu/index.php/ijsdir/article/view/147/152

and

Girardin, F., Calabrese, F., Dal Fiore, F. , Ratti, C., and Blat, J. (2008). Digital footprinting: Uncovering tourists with user-generated content. IEEE Pervasive Computing, 7(4): 36-43.
http://www.girardin.org/fabien/publications/girardin_ieee_user_generated_final.pdf

Visitor experience at the Louvre

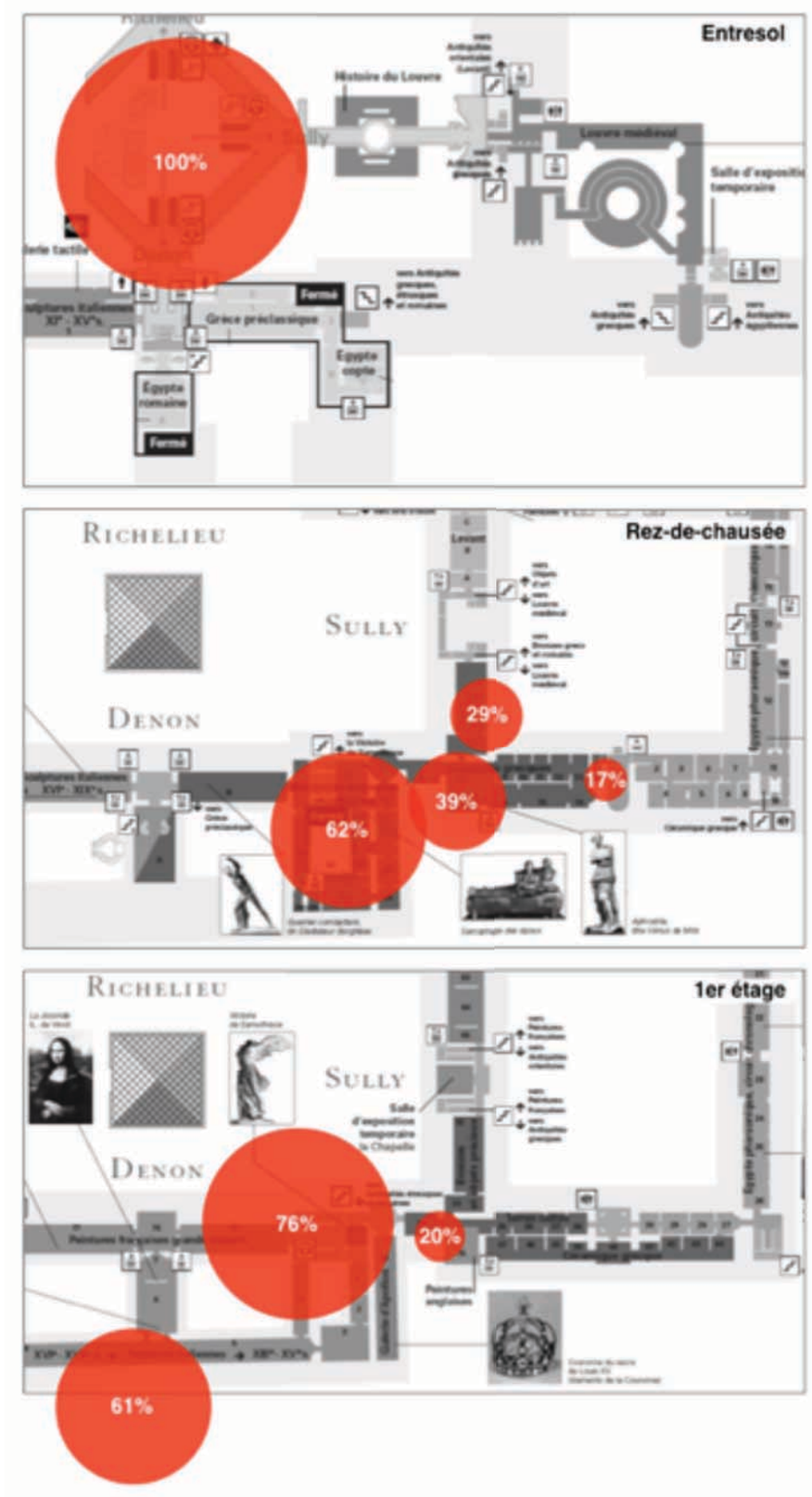
Another project that I am currently leading at Lift Lab brings us to the Louvre, by far the most visited museum in the the world with 8.5 million visitors (in 2009). In this context of “cultural enthusiasm” the museum witnesses levels of congestion, which, beyond a certain threshold can be described as hyper-congestion. This phenomenon has direct negative consequences on the quality of the visitor experience as well as on the organization and management of the Museum (e.g. increased stress level of the surveillance staff).

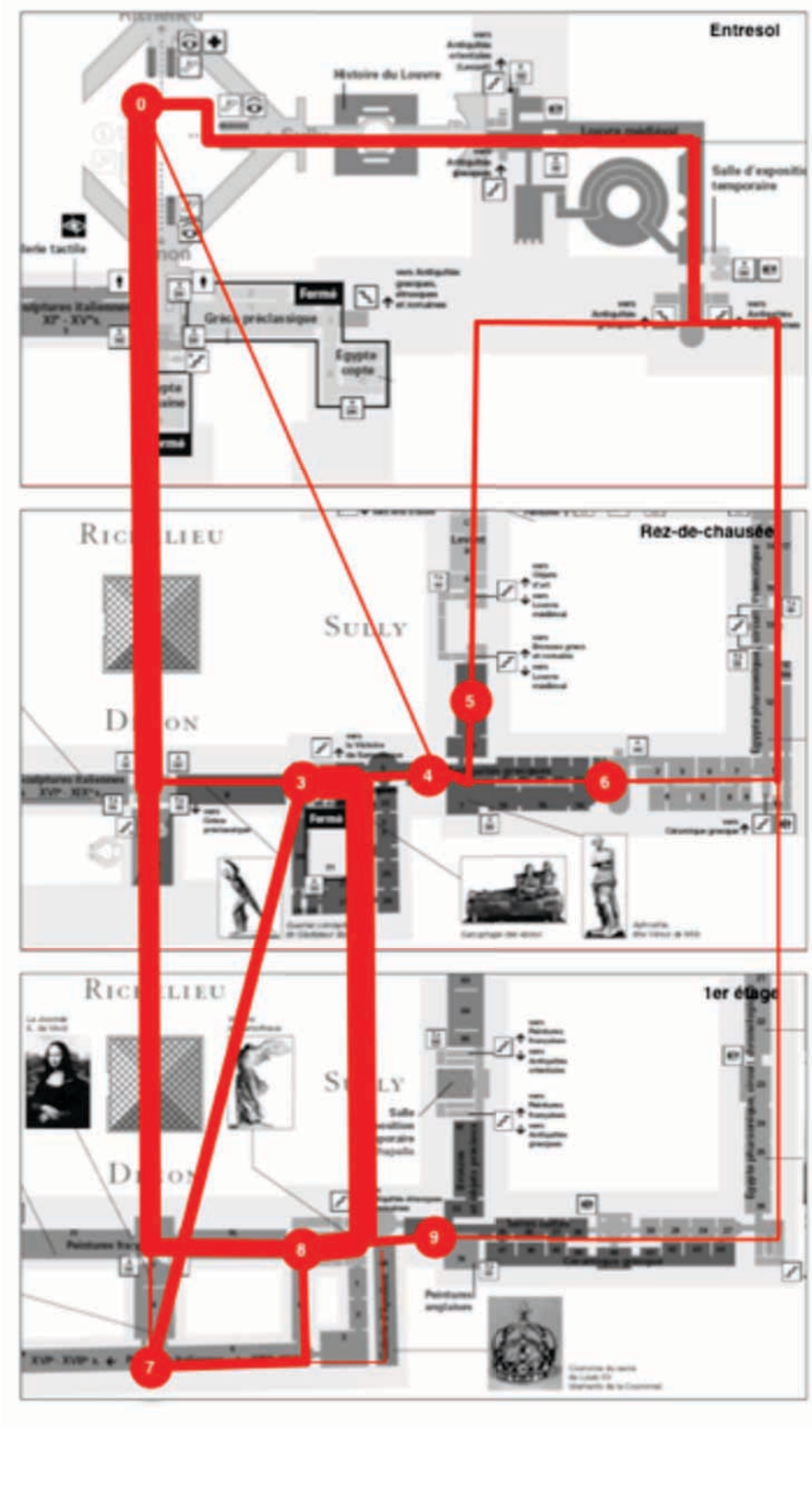
xxx employees .... xxx visitors (40’000/day at peak time)

Support space management strategies

We provide the Louvre with empirical and longitudinal indications of hyper-congestion that feed strategies to manage the space and ensure a good visitor experience. In collaboration with BitCarrier http://www.bitcarrier.com/ we perform audits on the presence of mobile phones in key areas of the museum at key moment of the year.

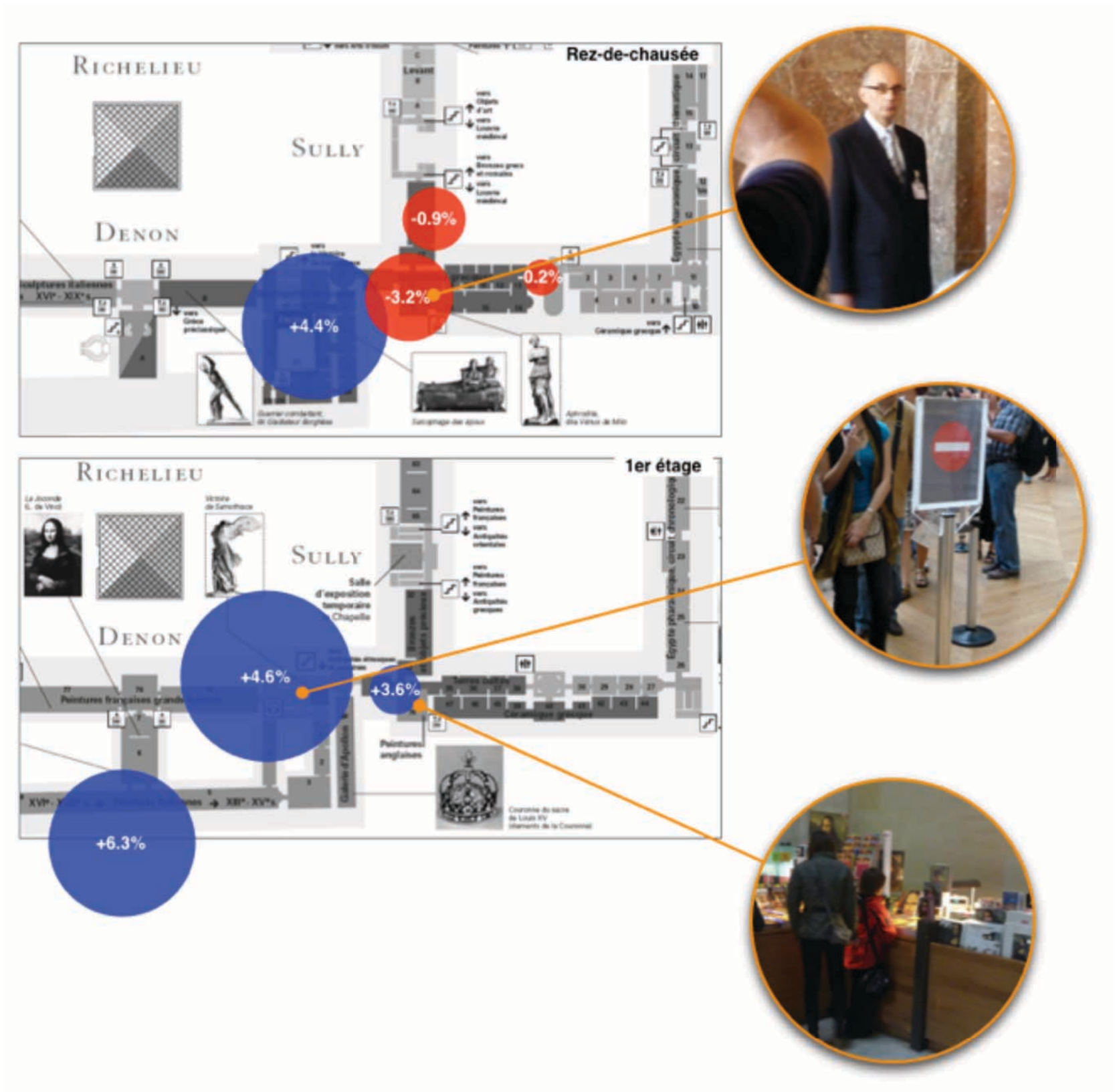
Measuring occupancy levels and flows





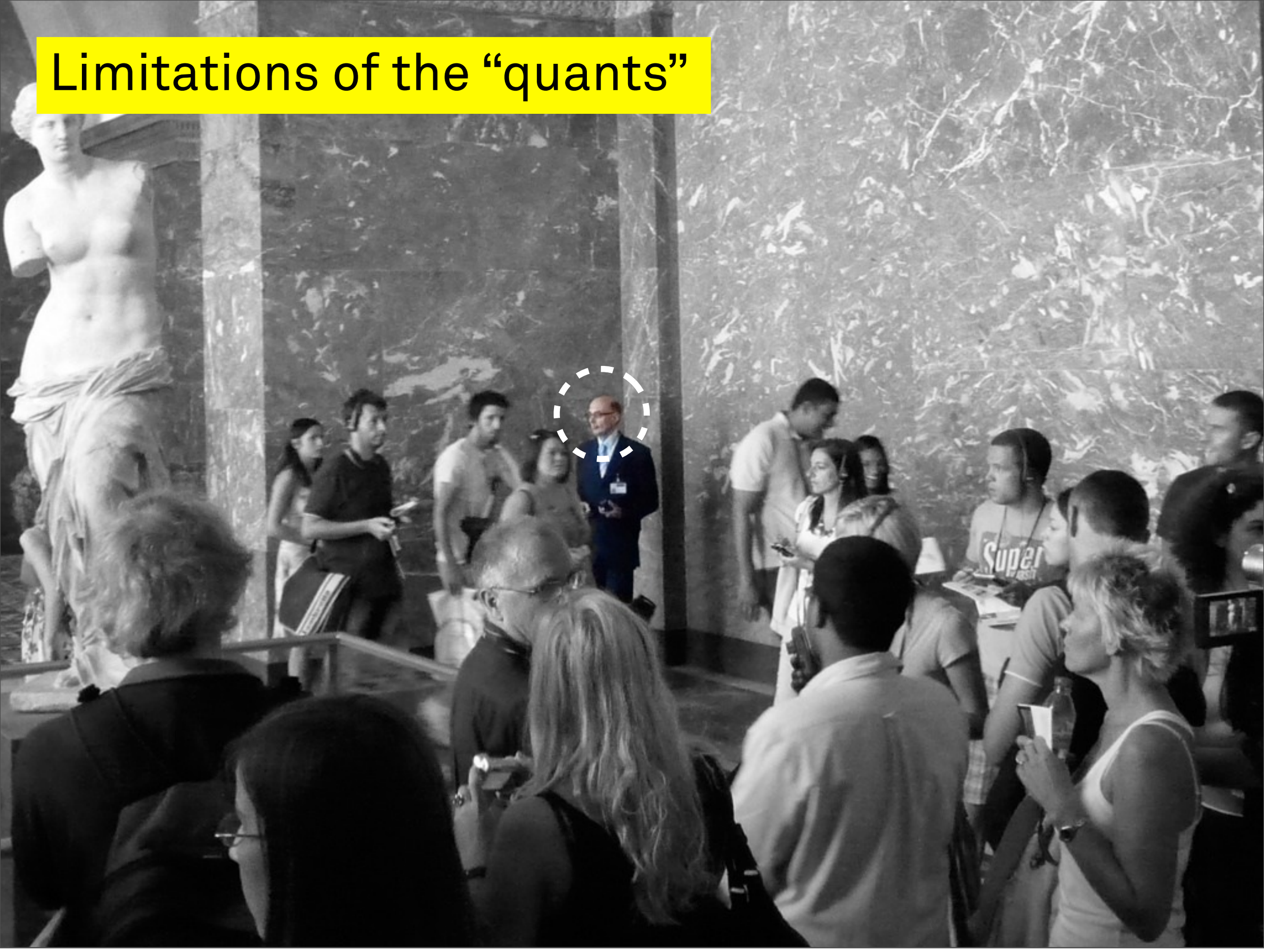
From the data of the aggregate presence of mobile phones we can generate raw material on the evolution of occupancy levels and flows within the museum (here obfuscated), particularly useful to measure the impact of certain strategies (such as the open doors of the first Sunday of each month).

Hyper-congestion indicator



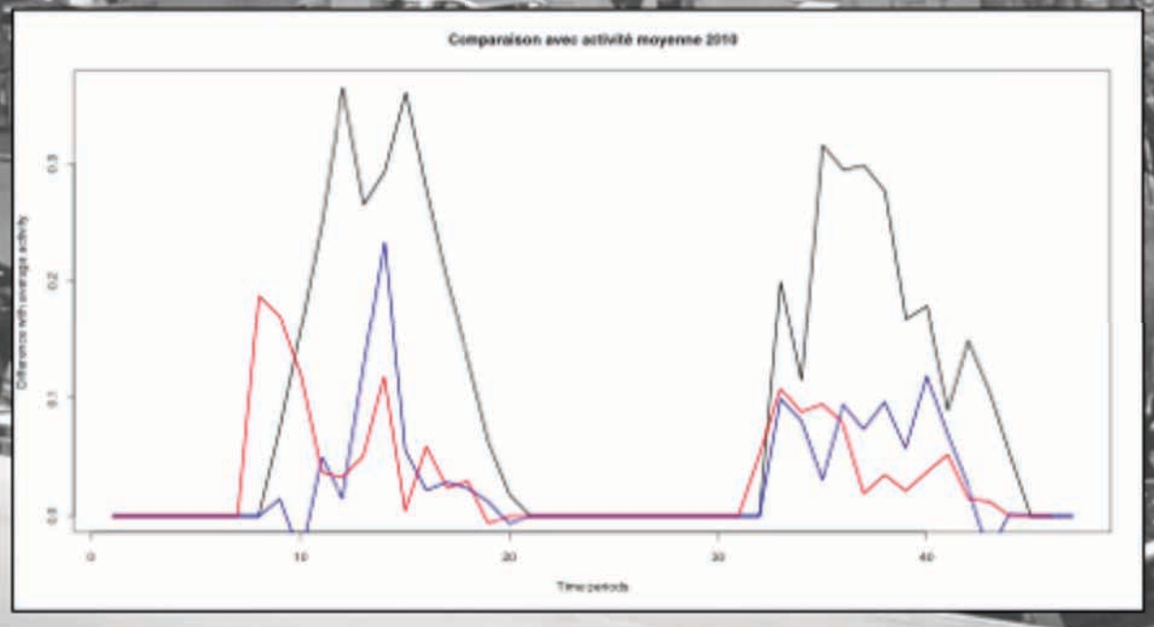
But particularly, this raw information is crucial to detect hyper-congestion situations (here obfuscated) to aliment particular strategies that aim at dissolving the presence of visitors. For instance. the surveillance team can temporarily redirects certain flows or use signs to close accesses. In a more indirect fashion, the museum thoroughly studies the location of information desks in less popular areas.

Limitations of the “quants”



Of course, 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, institutions and cities. The network and quantitative data we collect are not the sole sources of evidences to aliment the design of the space and its strategies. They are often incomplete to fully grasp a context and dynamics in a space. Qualitative observations on the field can prove very valuable. For instance, the surveillance staff is indispensable to help qualify the flows we measure at the Louvre.

Opportunities: Drilling as part of a space value



Finally, I would like to conclude this talk discussing a couple of opportunities from the exploitation of network data.

First, the capabilities to measure occupancy levels and flows has an impact of the value of a space. For instance, we are collaborating with the Geneva Palexpo exhibition center to help them measure the “hotspots” of exhibitions and conference. This information and its metrics is becoming an integral part of the commercial offer of Palexpo to rent its space.

Opportunities: New urban info providers

Second, even if we talk a lot about opening public data, other actors already offer the analysis of their data as part of their core service.

For instance, in Spain, Idealist has become a new type of urban information provider http://www.idealista.com/pagina/informacion. Idealist is a real estate portal to look for, rent or sell apartments/house. It is only recently that, based on the content and interactions on their website, they have started to offer almost instantaneous information on the evolution of the real estate market at a neighborhood scale. In the past, this information would have taken a census bureau long administrative procedures to produces unaccessible information on prices and level of interest in the housing market.

.

Opportunities: Citizens tweak urban systems


Inicio > Ciudadanos

Un joven barcelonés idea un algoritmo para mejorar el bicing

El cálculo permite fijar las rutas ideales que han de seguir las furgonetas para redistribuir las bicis con rapidez | La fórmula se ha aplicado ya con éxito a un centenar de paradas del servicio público de bicicletas

★★★★★ 0 votos | 42 comentarios

ÓSCAR MUÑOZ | Barcelona | 14/03/2009 | Actualizada a las 03:31h | Ciudadanos

El bicing esconde un complejo problema matemático, un rompecabezas numerológico que acaba de resolverse. Desde que se puso en marcha este servicio público de bicicletas en Barcelona, ahora hace dos años, y a medida que crecía, han ido multiplicándose las anomalías, muchas de ellas como consecuencia de la gran afluencia de usuarios que ha recibido y que ha desbordado todas las previsiones. Las soluciones se han ido aplicando sobre la marcha, sin tener un modelo previo de referencia. Sus

Finally, the presence of network data offers citizens new opportunities to appropriate and tweak urban systems.

For instance, in my hometown of Barcelona, the bike-sharing system Bicing suffers from malfunctions due to a poor distributions the bikes. After observing the situation a young student exploited the publicly available data on the bicing network to developed an algorithm that avoids both empty and full racks.

See http://liftlab.com/think/fabien/2009/03/14/citizens-to-improve-bicing/

I believe the next session will extensively discuss this topic, so I will not discuss it further.

Merci!

fabien@liftlab.com

I hope that this talk has helped you further understand the opportunities in exploiting network data as material to shape your or your city's strategies.

Thank you.

http://liftlab.com/think/fabien/