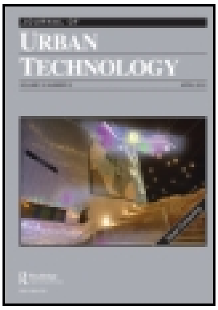

Journal of Urban Technology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cjut20

Enabling the Real-Time City: LIVE Singapore!

Kristian Kloeckl , Oliver Senn & Carlo Ratti

Published online: 05 Oct 2012.

Journal of Urban Technology, Vol. 19, No. 2, 89–112, April 2012

Enabling the Real-Time City: LIVE Singapore!

Kristian Kloeckl, Oliver Senn, and Carlo Ratti

ABSTRACT *The increasing pervasiveness of urban systems and networks utilizing digital technologies for their operation generates enormous amounts of digital traces capable of reflecting in real-time how people make use of space and infrastructures in the city. This is not only transforming how we study, design, and manage cities but opens up new possibilities for tools that give people access to up-to-date information about urban dynamics, allowing them to take decisions that are more in sync with their environment. This paper documents the ongoing LIVE Singapore! project which explores the development of an open platform for the collection, elaboration and distribution of a large and growing number of different kinds of real-time data that originate in a city. Inspired by recent data.gov initiatives, the platform is structured to become itself a tool for developer communities, allowing them to analyze data and write applications that create links between a city's different real-time data streams, offering new insights and services to citizens. Being a compact island based city-state metropolis, Singapore offers a unique context for this study. This paper addresses the value of stream data for city planning and management as well as modalities to give citizens meaningful access to large amounts of data capable of informing their decisions. We describe the technology context within which this project is framed, illustrate the requirements and the architecture of the open real-time data platform to serve as a base for programming the city, and finally we present and discuss the first platform prototype (using real-world data from operators of cellphone networks, taxi fleet, public transport, sea port, airport, and others). Based on this prototype a public showcasing of the project was staged in April 2011 at the Singapore Art Museum and the visual data analytics generated are illustrated in the paper. Finally, we draw some conclusions of technical as well as organizational nature regarding the challenges we faced when working in new ways with real-world, real-time data streams in an urban context that will help inform further development of our as well as of related projects in progressing in disclosing the potential of the wealth of digital data generated by urban systems, networks, and infrastructures.*

KEYWORDS *Real-Time City; Urban Data Platform; Singapore; Open Data; Stream Data; Feedback Loop; Urban System Dynamics; Urban Management Systems*

Introduction

Think of an encounter with a friend you ran into on the street. You both stop and exchange information—greetings, personal anecdotes, the weather, etc. In telling

Correspondence Address: K. Kloeckl, SENSEable City Laboratory/Lead, Real-Time City Group, MIT 10-485 77 Massachusetts Avenue, Cambridge, 02139 United States. E-mail: kloeckl@mit.edu

1063-0732 Print/1466-1853 Online.

http://dx.doi.org/10.1080/10630732.2012.698068

this story you formulate the phrases on the basis of your knowledge of the subject, as well as your knowledge of your friend's background. When he frowns his eyebrows, you understand that he disagrees; his gesture informed you he wanted to intervene and respond. You respond with a pause to give him an opportunity to contribute to the discussion. You note that your friend is getting impatient to finish the discussion when he looks at his watch. He, in turn, sees that you have noticed his impatience and explains he has an appointment to go to. All this interaction happens—in real-time. Your senses are capable of understanding this dynamic of interaction in a very effective way when the signals necessary for such an interaction can be perceived and when the signals are strong enough to be picked up.

Real-time interaction between people and these dynamics, instead, become difficult when distances between the person and the relevant phenomena grow (think of the size of a city) or when signals are of a kind that human senses have difficulty picking up or interpreting (think of pollutants contained in the air you breath). Now, consider the potential of allowing people's decisions to be informed by those phenomena, rendering dynamics on an urban scale as efficient and effective as that discussion between you and your friend.

Over the past years, urban space has been pervaded by networks and systems that generate digital bits. These bits are closely related to human activity, both informing human actions as well as reflecting their effects. Accessing the Internet on a laptop from a café locates the user through the WiFi antenna, using a cell phone registers at the nearest antenna tower, an online search or order of a product online establishes countless histories describing your preferences and interests, using a public transport smart card generates records on the origin and destination patterns of your trip, these, to name but a few examples, create a digital trace of your activity. This generation of bits happens incidentally, being generated and memorized during the daily acts of a city's life, and as such has the potential to reveal the complex facets of the city and its inhabitants in detail and in real-time (Rojas, Kloeckl, and Ratti, 2008). The increasing pervasiveness of urban systems and networks using digital technologies generates enormous numbers of digital traces capable of reflecting, in real-time, how people make use of space and infrastructures in the city. This is not only transforming how we study, design, and manage cities but opens up new possibilities for tools that give people access to up-to-date information about urban dynamics, allowing individuals to make decisions that are more in sync with their environment.

However, most digitally managed urban systems generate data for the purpose of their own operations and do not make those data directly accessible to other urban systems or the public. As a result, digital bits describing human activity in urban space related to these systems are accurately described by the generated data but locked within each of these specific domains. In the LIVE Singapore! Project, we are now exploring the development of an open platform for the collection, combination, and distribution of large numbers of a city's real-time data, allowing developer communities to join in creating applications that turn these data streams into meaningful and beneficial tools for people to make use of in their cities.

We have articulated this article in the following structure:

- LIVE Singapore!—The Project
  - Singapore: an ideal real-time city?
  - open data initiatives and platforms

  - mash-ups and combining data
  - an open platform project
  - users creating connections among data
  - feedback loops
  - system pictures
- Architecture for an Urban Real-Time Data Platform
  - data transport requirements
  - platform architecture
  - aspects of scalability and flexibility
  - data processing functionality
  - a first implementation and evaluation of a platform prototype
  - aspects of data format and semantics
  - LIVE Singapore! Exhibiting An Urban Demo
  - Critical Issues
  - Where Are We and What Is Next?

LIVE Singapore!—The Project

The SENSEable City Lab, housed at the Massachusetts Institute of Technology, has over the past years created partnerships with entities that manage large urban systems whose normal operation generates large data streams, and we have developed innovative ways of interpreting these and inferring urban activity as it happens in real-time. The Real-Time Rome project (Calabrese et al., 2007) analyzed data streams originating from Italy's largest cellphone network and from GPS locations of Rome's public transport bus fleet and its taxi fleet. The project's objective was to explore how using real-time data could allow for the better planning and managing of cities, with the main audience being planners and architects. A year later, Real-time Rome was succeeded by the WikiCity project (Calabrese et al., 2007) in which our work with real-time data was taken further, exploring a way in which combinations of data streams could be turned into effective tools for citizens to support the decision making process when following their goals and activities in the city.

The LIVE Singapore! project builds on these initiatives and takes them a step further. The project is not geared to any limited number of applications of real-time data in the city nor does it single out a few data streams; rather it develops a platform that is made accessible to developers in the city and beyond to share in making novel combinations and uses of a growing number of available urban data streams.

Over the past years significant work has been carried out in the domain of urban sensing such as that described in the CitySense initiative (Murty et al., 2008; Paulsen and Riegger, 2006). In these cases, sensor networks are deployed ad hoc, allowing for a high level of control of the forms of data generated by the system. Other projects, predominantly in the transportation domain (such as Massachusetts Bay Transportation Authority (MBTA) open data on real time bus locations), focus on making available existing data streams while remaining predominantly within one single type of data generated by one network.

In the LIVE Singapore! Project, we acknowledge that vast amounts of data are being generated in growing numbers of diverse urban systems as a kind of by-product of the specific system's operations. These data have the potential to reveal urban dynamics that go beyond the sole system of origin. Furthermore, such data often represent large parts of a city's population (consider the 143.6 percent cellphone penetration rate in Singapore as an example [IDA Singapore, 2012]), and in this way refers to a new urban context in which the daily reality of the city's inhabitants becomes the center of interest, as described by Zardini (2006). In this sense, LIVE Singapore! does not set out to solve any one specific urban issue as such but rather explores the development of new digital tools that enable radically new approaches to how urban issues can be addressed.

Singapore: An Ideal Real-Time City?

Singapore is in many ways an ideal setting for the exploration and experimentation of an open real-time data platform. Singapore is a city state and as such all territory is one city, both in terms of its dynamics as well as in terms of political administration—this is of great help in identifying and engaging with relevant partners for collaborations. Singapore's 5,076,700 inhabitants (2010 Census) live in an area covering only 710.2 km^2 of surface area. It is an island connected with the Malaysian mainland via only two bridges for motorized traffic; all other passenger and freight transportation is managed either via sea or via air traffic, with Changi Airport being a significant hub in the southeast Asian region and Singapore's container port being the world's second largest container port as well as the largest transshipment container port in the world. In this context, Singapore is a highly developed city, with advanced infrastructures and networks and in a tech savvy population familiar with recent developments in electronic mobile devices. Additionally, Singaporeans are familiar with rapid radical change in their city structure considering the country's rapid development since its independence in 1965 both in economic as well as in infrastructural terms: Singapore's annual GDP at current market prices has increased from US$ 1,919 million in 1970 to US$ 259,823 million in 2010 (Singapore Department of Statistics, 2012), while the city's subway system, opened only in 1987, has now expanded to four lines, 84 stations and 130km (Singapore Land Transport, 2012).

For all the above Singapore provides an ideal context for exploring research and development aspects connected to an urban, real-time data platform. It being so ideally suited, the risk is rather that not all discoveries and developments will likely be directly replicable in other urban contexts.

Open Data Initiatives and Platforms

LIVE Singapore! is inspired by the recent developments in the field of open data initiatives. In May 2009 the U.S. government launched the first of what has since been referred to as data.gov (data.gov, 2011) initiatives in which it made non-sensitive historic data sets publicly available in order to increase the "ability of the public to easily find, download, and use datasets that are generated and held by the Federal Government" (data gov, 2011). In this dynamic of giving data collected from the people back to the people, some of the goals are to spur economic, scientific, and educational innovation, and to promote citizen participation, as well as to drive down government costs by inviting a larger public community to actively participate in exploring how new value can be created from existing sets of data and their novel combinations. Since its start, the initiative has been replicated in various other countries (the United Kingdom) and cities (London and San Francisco) and multiple applications have been created based upon the data sets made available through these platforms.

Not limited to public data sources, platform projects such as Pachube (Pachube, 2011) have tackled the challenge of sharing sensor data over the web ("data brokerage"). The project was developed as an infrastructure for the Internet of Things that allows individuals and organizations/companies to make their sensor data available online. The platform uses a RESTful, HTTP-based API for the interaction among the sensors, the platform, and the clients which makes it easy to publish and consume sensor data.

Two different ways are supported for uploading data. The sensor can directly "push" data by issuing a HTTP PUT request along with the data. Alternatively, Pachube can also poll data from the sensor in 15-minute intervals or when a user is requesting data (if the sensor is supporting this mode). Pachube primarily focuses on supporting sensors with low sampling rates, since both mechanisms to upload data are based on HTTP requests and, therefore, are not suited for high-rate data streams.

Apart from acting as a data exchange platform between sensors and clients, Pachube does not provide any data processing functionality such as aggregation, advanced filtering, or combination of data streams.

Mashups and Combining Data

In a Web context, Mashup applications combine two or more sources of data to create a certain functionality or to simply derive new data sets. Mashup tools help in acquiring, combining, filtering, and aggregating the data. Furthermore, often these tools allow users to easily build (i.e., without programming skills) mashups and visualize or share the outcome of the mashup process.

One example of an online mashup tool is Yahoo Pipes (Yahoo Pipes, 2011). In analogy to Unix pipes, the idea is to apply several simple operators after each other to one or more data sources in order to extract the data that is needed. The tool supports several mechanisms to fetch data such as reading from CSV files, accessing RSS and Atom feeds or getting images from Flickr (an image hosting website). Yahoo Pipes does, however, not offer mechanisms to fetch data from data streams; it only polls the data sources in certain time intervals.

Once the data is loaded, different operators can be used to modify and combine data from different sources. There are separate operators that filter, sort, and aggregate data. To build the mashups, users employ a visual editor rather than a query or scripting language. After the mashup is finished, it can be viewed and reused by other users.

The functionality of Yahoo Pipes allows users to quickly combine various static data sets. The goal in LIVE Singapore is to have similar functionality but for real-time data streams. This increases the complexity of the operators significantly since data streams are infinite streams of messages but most operators (say the sort operator) rely on a limited set of messages with which to work.

An Open Platform Project

The LIVE Singapore! project leverages all of the above described dynamics. It is a flexible and scalable open platform structure enabling the collection, combination, as well as the distribution of large and growing numbers of real-time data streams originating from a city. Inspired by the data.gov initiatives, LIVE Singapore! is not a limited set of specific applications as such but an enabling platform on top of which developer communities, internal as well as external to the project group, can subsequently build applications that employ combinations of the data feeds available through the platform. In this way, instead of limiting the application development to a restrained group, we trigger a dynamic of crowd sourcing of the actual ideas created and the development of applications that become possible for the city, firmly believing in the creative potential of cities and their inhabitants. As in the case of the data.gov initiatives and their historic datasets, also in this project's context, letting citizens, city planners, companies, authorities actively work with live data streams has the objective of: fostering crowd sourcing, multiplying the creative potential, and ultimately setting the base for an emerging economy based on new ways to extract value from data. For this reason, the LIVE Singapore! platform aims at becoming somewhat an ecosystem with toolboxes that enable developers to write applications that turn urban real-time data into useful information for citizens.

While working with real-time data from any one single technology network has shown interesting results in recent years, it is becoming increasingly clear how the combination of very different types of data can bring about even richer insights into how cities work and how people make use of space. The data that we are considering for this project can be divided into three groups: data as a "by-product" from existing networks, data collected with tags or sensors, and data actively shared by people.

Dealing actively with a multitude of data that originate from very differently structured networks brings one main attention to the conception of connector elements as outlined in the next section. While we are actively developing code to include data streams in this first prototype phase, we will be moving towards providing indications as to how external data providers can provide data to the platform in a more standardized way. It is clear that a platform such as LIVE Singapore! needs to tackle and resolve the seemingly conflicting requirements of providing connectors that require as little data processing and conversion on the provider side as possible (in order not to discourage participation), yet receives data streams that are as coherently structured and described as possible for a maximum of interoperability and the possibility of combining different data streams. Striking a balance is a challenge well known in the study of system change:

> If connections are too few, networks are frozen and no change occurs, and if connections are too many, there is no stability and networks remain chaotic. Along the margin between too little and too much connection, 'the spontaneous emergence of self-sustaining webs' occurs,

says Mark Taylor (Taylor, 2001: 148), quoting biologist Stuart Kauffman.

By creating a platform that enables previously impossible or hard-to-achieve links among diverse data streams, LIVE Singapore increases the complexity of the urban system, since "complexity is directly related to connectivity. As connections or interconnections proliferate, complexity expands and, correlatively, information increases" (Taylor, 2001:139). And this increase of complexity, of connections among elements which in this case are data streams originating from very diverse sources in the city, is what the project ultimately aims at: creating a platform for developers to connect multiple and diverse data streams, generating in this way unprecedented insights into a city's dynamics as well as developing tools and applications for citizens to navigate through them and act upon them in a way that is constructive for both their individual as well as collective goals.

For this reason, another of the platform's declared requirements that orients development is to provide easy access and easy programmability in order to capture as large as possible an active developer audience, interested and able to program applications that valorize the available data streams.

Users Creating Connections among Data

Currently, much of the value that is being attributed to many online platform projects is the data that these platforms accumulate and which give them deep insights into their user's activities and preferences. The situation for LIVE Singapore! is of a fundamentally different nature: This project deals with stream data and as such does not store any data within the platform—it is always the data provider that has control over what data is made available. We believe it would be difficult for any one operator to propose itself today to centrally host a wide variety of a city's data streams for difficulties related to data management performance but even more so for questions of data ownership, privacy, and security. The platform is, therefore, structured to make data streams available in a coherent form but not to store these data. Historic data can be made available in much the same way as the live streams, being hosted on the providers' site and made available through connectors via the platform.

LIVE Singapore! in this way sees the value of data not in it's centralized accumulation but focuses on keeping track of the connections made between real-time data streams by developers and users creating applications. While the semantics of data streams is a challenge, even more so is the semantics of data connections: what is the value and meaning of the combination of any group of data streams? Our approach lets the users make and describe these connections while saving this information within the platform in order to gradually grow an internal knowledge about the semantics of stream connections to inform the platform's growth.

Feedback Loops

We have indicated above how people's actions in cities translate into data in multiple technology systems. Providing information on urban dynamics (gained through the collection and analysis of those data) to people when they are in the process of making decisions, creates new and effective feedback loops to the city system:

> In an information-feedback system, conditions are converted to information that is a basis for decisions that control action to alter the surrounding conditions. The cycle is continuous. We cannot properly speak of any beginning or end of the chain. It is a closed loop. (Forrester, 1961: 61)

While it is certainly still the case that in such a system "it is always the presently available information about the past which is being used as a basis for deciding future action" (Forrester, 1961: 15), working with real-time data today as described does move that past significantly closer to the present and can reduce delay, allowing for more timely and potentially effective decisions and actions. While Forrester points out that delays in systems do not necessarily have a negative role but can also have a stabilizing one, the feasibility of shorter delays does open up new possibilities and increases the flexibility with which delay time in feedback loops in urban dynamics can be put to use for more effective employment of resources and energy in the city.

Important questions emerge regarding the impact that such new real-time feedback loops will have on a city and on its dynamics—will they break the stability of the system? Or instead, might they create stability where there is none yet? Since Forrester points out that

> there is strong indication that among real systems of importance some are unstable in the usual mathematical sense. They do not tend toward a state of static equilibrium (even in the absence of randomness or external disturbance). They are unstable, tending toward increasing amplitudes of oscillation that are contained by a continuously shifting balance of forces among the system nonlinearities. (Forrester, 1961:66)

We might take some cues from Donella Meadows when she points out that while a system's behavior is determined jointly by its parts, its inter-connections as well as its function or purpose, "changing inter-connections in a system can change it dramatically" (Meadows, 2008:17).

System Pictures

When dealing with real-time data from urban systems and networks, the SENSEable City Lab has always put as much emphasis on the research and development of adequate software platform architectures, the data analytics, as well as the communication of the information contained in the data collected and studied in the urban context. Visualizations have proven particularly useful in this undertaking. As Meadows describes so well,

> words and sentences must, by necessity, come only one at a time in linear, logical order. Systems happen all at once . . . To discuss them properly, it is necessary somehow to use a language that shares some of the same properties as the phenomena under discussion. Pictures work for this language better than words, because you can see all the parts of a picture at once. (Meadows, 2008:5)

This characteristic of being able to grasp the whole picture at once is also the reason why control rooms dealing with large amounts of data do in fact put an emphasis on diagrammatic representation rather than leaving the data in their numeric form or translating them into verbal expressions. A second reason for the strength of data visualizations is the additional processing power that can be tapped into on the spectator's side. Seeing an animated visualization of urban real-time data involves the spectator's processing power. He will bring in his prior knowledge and experience in analyzing and filtering the information in the process of observation, similar to a person standing on top of a high-rise building to observe the traffic situation before deciding in which way to drive—in this case the spectator brings in his knowledge of past dynamics and integrates them with what he sees in real-time.

Architecture for an Urban Real-Time Data Platform

From the above illustration of the overall project vision and concept, we have made an attempt to derive a platform architecture that is distinguished by two main parts: a messaging part (transporting data from a provider to a set of consumers) and a usage part. The platform is designed to handle data in the form of (real-time) data streams, which consist of an infinite sequence of messages with each message having a number of data fields. The second part of the functionality consists in data consumers being able to manipulate and process the data.

Data Transport Requirements

LIVE Singapore! has the unusual characteristic that it is not as such an application with a clearly definable use case scenario and context. It is a platform that aims at enabling a self-sustaining application development activity. We decided not to take example applications as orientation help in order to avoid slipping into customizing the platform towards a limited number of specific applications. Some cornerstones, however, must be and can be identified. LIVE Singapore! will make data streams available to the public. In the project's first phase, seven data sets and streams have been provided by government agencies, institutions and companies. The platform will also support data streams from individual devices (like smart phones) in the future. This implies that our platform needs to be able to handle a large number of data streams. For development purposes, we now assume hundreds of data streams, realizing that this number could go into the thousands.

The data streams from data providers vary widely in terms of data rates. A large telecom company partnering with us provides us with a data stream with a rate of around 300 messages per second. At the same time, the platform also has to be able to handle data streams with only a few messages per second. This is especially true if we consider individuals contributing data from their smart phones. Considering a growing number of providers in the future means that the platform is required to ultimately handle tens of thousands of messages

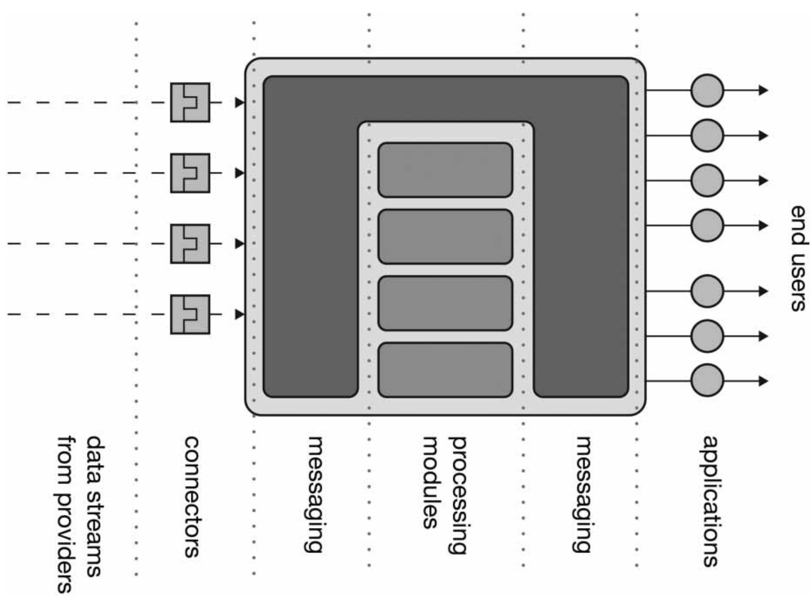

Figure 1: Platform architecture

per second in overall input load. On the output side, once the data of the platform is made publicly available, many people will be using the data (for now we assume several thousand) resulting in each data stream being distributed to a potentially large number of clients.

Taking into account the overall input load and the numbers of potential clients, and doing a quick approximation, we easily end up with up to one million messages per second. Since the initial load of the system will, however, be much smaller, it must be designed in a way to run on just a few servers initially but scale up near linearly with every server we add.

Platform Architecture

The overall architecture can best be describeSd by following messages of data streams through the platform. (See Figure 1)

Since each data provider tends to have its own data format, for data processing to become possible, the data needs to be converted into a common data format. The platform provides connectors to the data providers that allow the data to be converted into the platform data format and then forwarded to the platform (i.e., connecting to the platform messaging system).

The messaging system is then responsible for receiving the data from connectors and routing it to the client applications that need it. A client application can subscribe to individual data streams. After that, every message of that data stream is then pushed to the client.

The processing modules implement the data manipulation and processing features of the platform. They take one or several data streams as input (via the messaging system), do the processing/manipulation and output a modified data stream (to the messaging system). Client applications can then subscribe to the modified data stream in the same manner as they would for an unmodified data stream coming from a data provider (i.e., through the messaging system).

Aspects of Scalability and Flexibility

As mentioned in the requirements paragraph above, we expect that the platform will have to deal with a considerable number of messages (this causes CPU, memory, and network load). Furthermore, the processing modules impose additional load. In order to handle this load, the work has to be distributed to many machines.

While the system is designed to run on a few servers initially, it is also designed to scale up to hundreds of machines to accommodate a growing load. Our approach to achieve this scalability is that each part of the system (we call the parts components) can be run on separate machines if necessary. Each component can also be replicated on many machines if needed. Components are, for example, the individual processing modules and the parts of the messaging system.

We illustrate this by describing a common scenario: it is very likely that some particular data streams will be very popular with client applications. Many client applications will subscribe to these data streams and a single machine may not be able to distribute the messages to all the clients anymore. A component called Forwarder (See Figure 2) is then used to replicate the data stream to many machines.

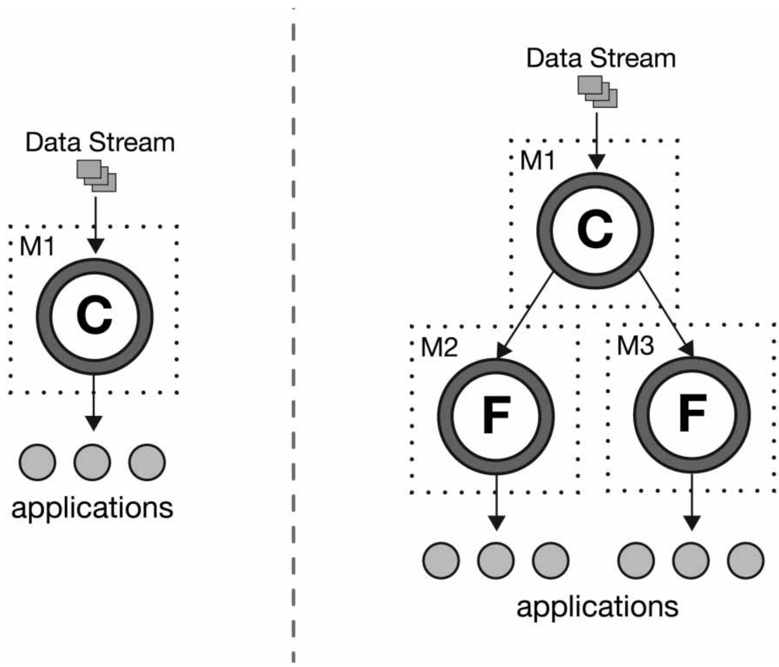

Figure 2: Forwarders replicating data stream on many machines

(In Figure 2, the data stream is replicated to two machines.) This distributes the load among two machines and enables us to support many client applications. The cost of this solution is an increased latency for messages since the messages have to be transported over one link more.

The concept works the same for other components of the system such as the processing modules. This very flexible way of running components on different machines and, therefore, distributing load works because all the components have a similar interface—they consume data from a number of data streams and output a data stream.

So far an administrator has to configure which components to run on which machines, but we plan to develop an automatic mechanism that starts and stops components depending on the current overall load of the platform and the load on individual machines.

Client applications subscribing to data-intensive data streams and/or using complex filter modules cause a heavy processing load on the platform. An advanced concept for limiting the load is to offload some of the computation to the client applications themselves. For example, the client application requesting to filter a data stream could run the filter module locally, thereby freeing platform resources.

The platform could even be outfitted with a cost model that balances computation costs and costs for transporting messages to clients. If, for example, two client applications execute the same filter over a data stream, it might be beneficial to run the filter on the platform to reduce the number of messages that need to be forwarded to the two client applications. The cost model could be made dynamic so that it considers the current platform load in terms of computation and networking load.

Data Processing Functionality

Besides transporting data from providers to consumers, LIVE Singapore! allows clients to easily process, manipulate, and visualize the data of data streams. Several modules have already been implemented in a prototype:

A Predicate Filter Module. Allows messages of data streams to be filtered. The user specifies an arbitrary complex predicate that is then evaluated for each message of a data stream (and the message is then either discarded or forwarded).

As an example, consider a data stream with a data field for humidity and temperature. The predicate could be temperature $>=10$ meaning that the filter would discard all messages where the temperature is below 10 degrees.

A Point-In-Polygon Filter Module. Often data streams contain geo-tagged information and many client applications may only be interested in data from a certain area. This module allows users to specify a polygon (using lat/long coordinates) and the filter then only forwards messages that have a location within that polygon.

A Truncating Module. This module allows users to remove unwanted data fields from data streams.

Our system can easily be extended with new modules. We plan to implement several modules in the future:

- A module to support "windowing" (well known from stream-processing engines). Windowing is needed, for example, when a user wants to calculate the average value of a data field for some messages. Since a data stream is an infinite sequence of messages, the user needs to be able to select a meaningful subset of the messages to calculate the average.
- A module that converts data from our internal data format to different well known formats (such as XML, CSV, JSON) requested by client applications.
- One very specific module that we would like to implement is a map-matching module that would be of great use for all transport-related data.

Since all the modules are implemented as components, they can all be linked together, allowing users to build processing chains out of simple operators. This idea is again very similar to stream-processing engines. The downside with stream-processing engines is that the operators (especially when it comes to windowing) are very complex and hard to understand and use. We plan to limit the number of modules and their functionality to make them easy to understand and use.

A second advantage of modules having data streams as outputs is that all client applications can potentially subscribe to, say, a filtered data stream created by another application. This allows users to build up a repository of useful modified data streams. Also, the platform will keep track of all the derived data streams so that we can add a mechanism that advertises popular derived data streams to users (since those are most likely meaningful ones).

A First Implementation and Evaluation of a Platform Prototype

In 2011, a first prototype of the platform was implemented. The messaging system was implemented using a low-level (TCP) messaging library called ZeroMQ

(ZeroMQ, 2011). The advantage of that library is that it allows us to implement a fast-messaging fabric both internally as well as between the platform and data providers/client applications.

The messaging system underwent some initial tests that showed that even a normal desktop machine is capable of distributing a data stream with a data rate of 1,000 messages per second to 300 clients in parallel with an average latency of around 1.3 seconds. This result shows that the overhead imposed by the messaging system is relatively low. However, many more tests will be necessary, especially to see whether scaling the messaging system works in a nearly linear manner (meaning twice the number of machines can handle almost twice the load).

Apart from the messaging part, two processing modules (predicate and point-in-polygon filters) were implemented. Also, a management console was provided allowing for the managing of the network of components that are used. Another feature of the prototype is that the new HTML 5 WebSocket protocol (WebSocket API, 2011) is supported. This means that messages of data streams can be pushed to a Web browser (or any other HTTP/HTML client). This is an important feature because, unlike the low-level TCP, it works for clients behind a firewall/NAT.

In the future, we will do a more extensive evaluation of the platform under real-world conditions and implement additional processing modules (as mentioned above). Furthermore, we plan to provide users with a toolkit that makes it easy to subscribe to data from the platform and to handle and visualize it.

Aspects of Data Format and Semantics

At the current state of the project, the data model for describing city entities and the city data is proposed to be defined using the standardized SensorML Observation and Measurement Language.

The semantic description of the city entities and all the sensory data sources is specified using the ontologies defined in the Web Ontology Language. A Domain Ontology connects all sensory data ontologies.

The semantic description of the data is decoupled from the data per se, which can instead be represented in JSON, to reduce file size and parsing overhead due to the unnecessary descriptive tags. Figure 3 shows the generation process of the data model.

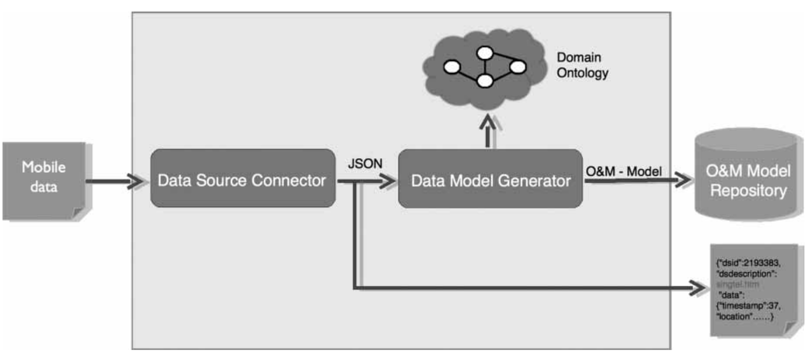

Figure 3: O&M data model generation process

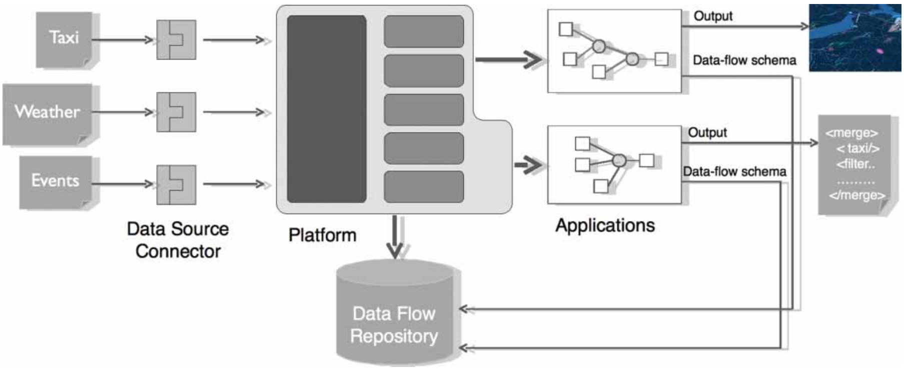

Figure 4: Central user-generated, data-flow process

Figure 5: Photos from the LIVE Singapore! exhibition that opened April 7, 2011 at the Singapore Art Museum (SAM)

Users can use the platform to search, browse, and integrate data streams based on their semantic descriptions, and create applications. The system overview is illustrated in Figure 4. Through the platform, applications can also be published and shared among users.

LIVE Singapore! Exhibiting an Urban Demo

LIVE Singapore! is all about making real-time data available to citizens and in this spirit an exhibition was organized in April 2011 at the Singapore Art Museum. The initiative showcased some initial steps and accomplishments of the project, giving a broad audience the opportunity to engage with data generated by their city. The LIVE Singapore! exhibition, staged in the mid-nineteenth-century former Saint Joseph's Institution building, now Art museum, comprised five projections and one LCD screen showing dynamic visualizations driven by data sets and live streams from the following data providers: (See Figure 5)

- *Changi Airport Group*, operator of Singapore's airport, which is a major hub in the southeast Asian region and the seventh busiest airport in Asia by passenger traffic in 2010 (Changi Airport Group, 2012). The data set consists of the number of actual boarded passengers on all flights arriving at and departing from Changi Airport.
- *ComfortDelgro*, one of the worlds largest transport companies operating also outside of Singapore where it operates a 16,000-vehicle taxi fleet. The data set provided consists of a GPS location, the status (occupied, unoccupied, booked…), and the speed information from all taxis at irregular intervals of about one minute on average.

- *NEA*, Singapore's National Environmental Agency provided a data set indicating the rainfall across the island in terms of intensity of precipitation, derived from rain water sensors as well as radar images.
- *PSA*, operator of the port of Singapore which is the second largest container port and largest transshipment container port worldwide. The real-time data stream details all containers arriving and departing, including origin and final destination ports as well as information related to vessels' actual arrival and departure times.
- *SingTel*, Singapore's prime telecommunication company and cellphone operator. The real-time data stream provides anonymized call detail records, cellphone network usage, and cell-ID data in 15-minute intervals.
- *SP Services*, the city's electricity provider provided an aggregated data set consisting in all electrical energy consumption by commercial and industrial clients (making up 70 percent of Singapore's total electricity consumption) for each of the city's 82 postal sectors in 30-minute intervals.
- *NUS Department of Geography*, provided a public data stream from a wind speed sensor.

The exhibition setting at the Singapore Art Museum was composed of the following six visualizations, two of which offered user interaction:

Isochronic Singapore

Based on travel data from 16,000 taxis in Singapore the fastest road traffic connections are calculated and allow for the user to select an origin location. Upon this selection, the map distorts Singapore's shape to reflect the time it takes to get from the selected origin to any other part of the island. In short, if traffic is slow

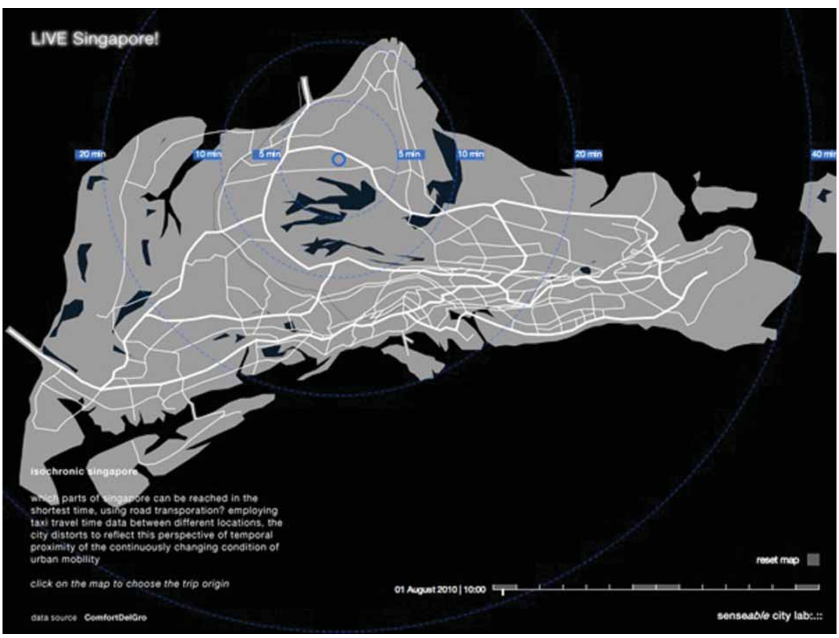

Figure 6

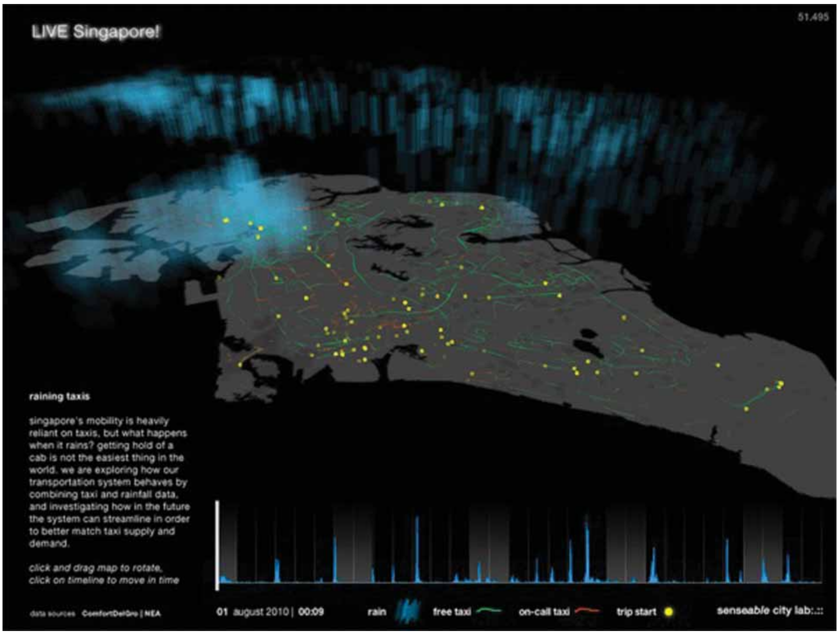

Figure 7

and travel time is long, the island grows larger while with fast traffic and shorter travel time, the island shrinks. Different areas distort differently reflecting the specific travel time to these destinations at a given time (See Figure 6).

Raining Taxis

Being located in a tropical climate, Singapore experiences frequent and often heavy rainfall of short duration that is limited to very constrained geographical areas with significant impact on transportation. This visualization brings together unoccupied taxi traces and taxi trip origins with the areas afflicted by rainfall, suggesting a multitude of possible user applications related to how to help improve supply-demand matching in high transportation demand areas caused by rainfall (See Figure 7).

Urban Heat Islands

As heat is generated by energy use, high energy consumption can translate into small local temperature rises of up to a few degrees (called man-made or anthropogenic heating). Combining data on the energy consumption of the city's different zones with the wind speed, local temperature rise can be estimated. In short, the higher the wind speed, the less local temperature rise is caused by local energy consumption due to dispersion. A potential future addition of actual measured urban temperatures will provide the basis for a future heat monitoring program (See Figure 8).

Formula One City

Large scale events disrupt a city's daily routines. Singapore's Formula One Grand Prix offers an excellent opportunity for observation. On top of a map of the city's

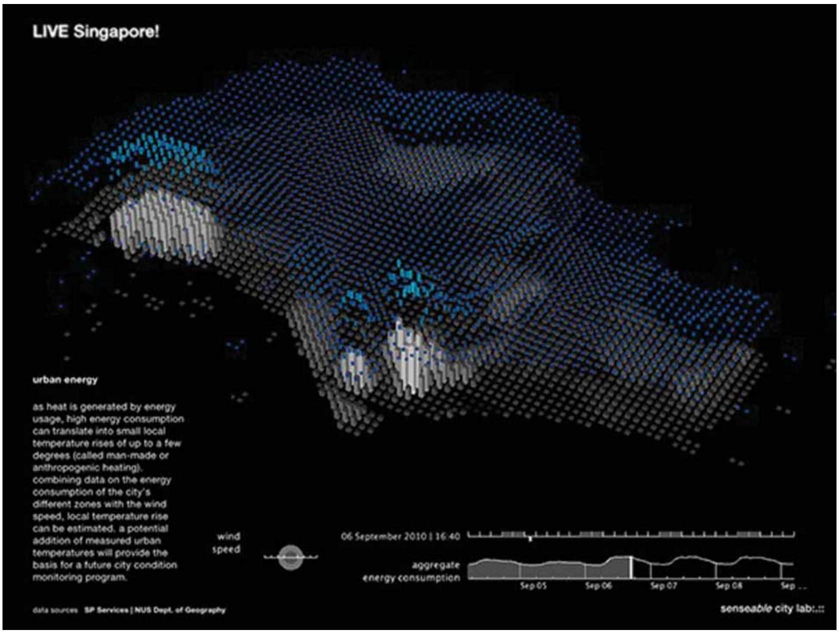

Figure 8

Figure 9

central district with the indicated F1 course (which in Singapore is located in the very center, composed of normal roads blocked for the event) visitors can see the network volume of SMS text messages being sent via the cellphone network antennas located in the area during a normal Sunday (on the left) and during the F1 racing Sunday in 2011. SMS text message traffic is more intense as the glow points increase in size and color intensity (See Figure 9).

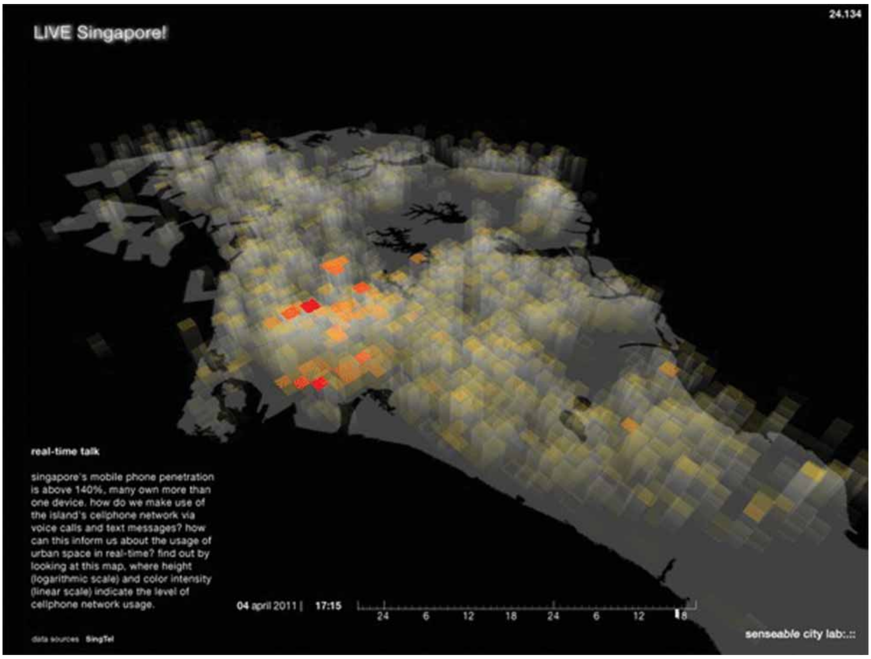

Figure 10

Real-Time Talk

Singapore's mobile phone penetration is above 140 percent, and visualizing in real-time where cellphone devices are located can give a good indication of the population distribution in urban space. The real-time data stream driving this visualization shows network call traffic intensity in terms of column height on a logarithmic scale and in terms of color intensity on a linear scale (See Figure 10).

Hub of the World

Singapore is the world's largest trans-shipment container port and one of the busiest airports in the world. Figure 11 conveys the city's role as a hub, conveying how the island is affected by the constant stream of people and goods passing through. The number of containers moving in and out from Singapore's port is represented by a line that links origin and destination of the containers and that varies in thickness according to the container number (See Figure 11).

Using the Data

The organization of the LIVE Singapore! exhibition had several purposes: Data shown in the visualizations were data generated in one way or another by people in the city. Sharing these data with the people of the city conveys the core mission of the open-platform research project. It also familiarizes and engages the city's public in critical discussion about opportunities and risks of working with urban real-time data, something we see as an important component for moving towards a real-time city. Working on the technical components of the

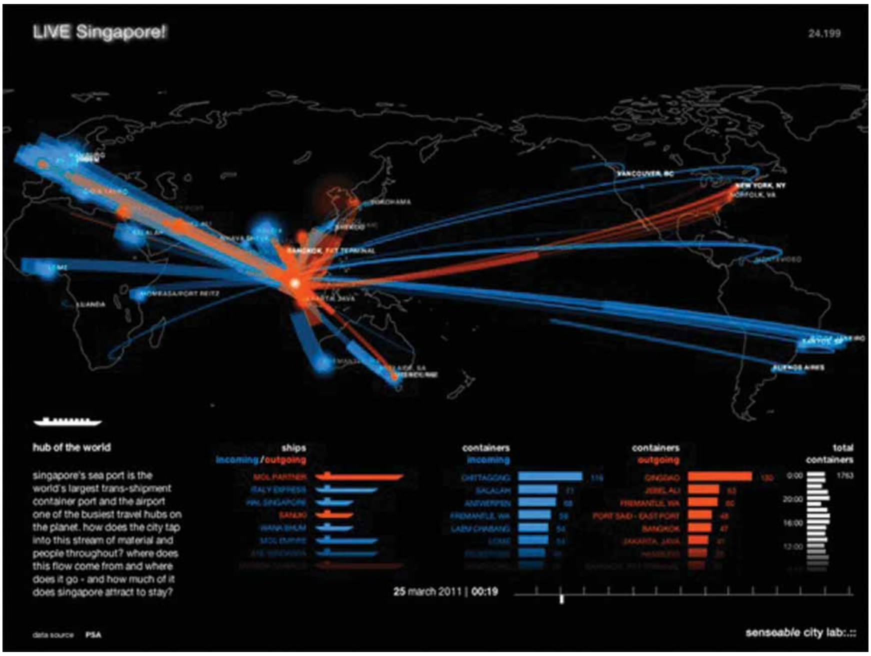

Figure 11

system has given us an exceptional opportunity to test components under real-world conditions, having to deal with a wide range of conditions and circumstances far from the theoretical concept—something certainly necessary in supporting with relevant research the development of technologies that will enable wide ranges of real-time data generated by the city's systems to be shared with its inhabitants.

Related to the domain of urban planning and management, the six real-time data visualizations described above showcase the potential for the development of applications using combinations of real-time data streams that already exist in cities. In particular, the combination of rain data and the location of free taxis in the context of a tropical climate suggests new opportunities to better accommodate high transportation demands caused by meteorological phenomena. The combination of location-specific data on electricity use and wind speed offers a contribution to the comprehensive domain of urban heat islands. And the use of taxi location data to visually represent estimations of road travel time allows for an instantaneous assessment of the road traffic situation in a specific city as well as visual comparisons among different cities.

Critical Issues

Recalling the scenario offered earlier—the description of a real-time interaction between two persons knowing each other and in a direct encounter—this scene contains a fundamental requirement indicating one of the main critical points of open data platforms such as the ones addressed in the LIVE Singapore! platform: privacy. The real-time interaction between the two figures will be the more effective the more they know about each other, be it in terms of prior knowledge or in terms of getting to know about each other as they interact. This allows each to correctly interpret the behavior of the other. In an open data platform this analogy brings up the critical notion of privacy in dealing with data collected in urban networks—data that are often directly linked to an identified citizen due to reasons of system functionality. As much as digital traces enable a system-wide awareness of urban dynamics that can constructively inform decision making in real-time, the inverse risk is what Bill Mitchell refers to as *electronic hyper-visibility*, expressing deep concern of such a reality if created without a "framework of vigilant, carefully debated, vigorously formulated and executed public policy" (Rojas, Valeri et al., 2008).

The issues of data privacy as well as data security need to be addressed in parallel with any project such as LIVE Singapore! and we have initiated an active discussion with policy makers in Singapore as well as at MIT.[1] As much as research on technical possibilities can inform policymaking, discussions on policies capable of expressing public opinion on what a society desires to do with data can and need to inform the research on the technical capabilities of a data-driven platform. While much of the public debate appears to be focused on restricting data use to ensure privacy protection, an alternative direction has been indicated by Weitzner et al. (2008) that rather proposes new forms of transparency on data use linked to clear policies on authorized uses. Over time, the issue of memory of user data has become critical when "remembering" becomes less costly than "forgetting" as Mayer-Schönberger (2011) illustrates, indicating new issues and forms of dealing with memory in the context of digital data.

At present, we have been handling data privacy within the LIVE Singapore! project by means of data aggregation and data anonymization, which have ensured concealing the identities of individuals associated with digital data streams. While effective, these approaches need to be subject to further examination regarding aspects such as the modality of aggregation (i.e., how fine or coarse does an aggregation need to be).

While the LIVE Singapore! project must, of course, be sensitive to the privacy of urban data platforms, when it is fully functioning, responsibility for the privacy of data will be more appropriately handled by the providers of the data. Considering hundreds if not thousands of data stream contributors, it is impracticable to think an open platform could handle privacy more efficiently than providers who should provide data of a kind and in a format that respects the privacy of individuals, whose information is released.

Another issue to be addressed as the platform develops will be the emergence of re-enforcing feedback in urban dynamics potentially caused by combinations of real-time data made available to citizens. When early in-car navigation devices rerouted drivers to avoid traffic jams, one consequence were secondary traffic jams on parallel roads. Will people change their decisions based on real-time information made available to them? This will be directly influenced by the way data are made available, both in terms of aspects such as delays and aggregations or combinations of data as well as the designs of interfaces by which people access that data.

A second kind of re-enforcing feedback possibly created by real-time data platforms stems from unprecedented datasets made public and combined with other data to lead to novel insights formerly unavailable to the public. An example for this would be data on the closing and opening of electricity contracts as a reflection of the behavior of the real estate market of specific neighborhoods. The public accessibility of such data could easily lead to a re-enforcement of a perceived trend and be turned into a much larger trend without any other information being available than the description of an initial state.

Aspects such as these fundamentally emerge whenever radical innovation of technology, policy, or of another kind changes the rules of how elements of the urban system connect. They need, however, be observed and studied with care in order to find their new place within that system and its modalities of interaction.

Where Are We and What Is Next?

The LIVE Singapore! project is a five-year research initiative started in mid 2010 and has reached the following state: a platform prototype has been implemented as outlined above, allowing at present the development of applications on top of the platform that is at this stage operating internally.

In April 2011, the LIVE Singapore! exhibition was staged at the Singapore Art Museum in which we showcased six applications developed within the LIVE Singapore! framework that combined multiple real-time streams, making them accessible to the public by means of visualizations (some of which allow user interaction) of the data that gave insights into the city's dynamics as they happened and demonstrated the usefulness of real-time data in concrete urban contexts. These applications were realized with data streams coming from an initial group of data sharing partners.

On the basis of these developments several conclusions can already be drawn that are informing the next steps of the project's development:

- It emerged in the collaboration with all partners sharing data with the platform that they had to make a major effort to prepare the data to be shared. The effort is manifold and some of the core points include: ensuring data privacy, extrapolating relevant and non-sensitive fields as well as structuring the data in a coherent format. This confirms the common idea of the existence of "data silos"—highly developed data systems where data are difficult to combine—these systems have not been set up with the purpose of sharing operational data with an external entity such as an open urban data platform. The process of interacting with our initial partners will substantially inform the formulation of standard procedures, policies, and data structures for the next phase of the platform, keeping in mind the above outlined contrasting objectives of a high level of data coherence as well as reduced requirements of data preparation by the providers. On the positive side, however, is the high level of commitment by the data-sharing partners who are motivated to address these issues, indicating an awareness of the overall benefit of an urban data platform such as LIVE Singapore! In some cases, the project has been taken by partners as a stimulus to address internal data incoherences and incompatibilities—an effort that will prove useful both for work done with data within the single network as well as preparing it for combined uses such as within the LIVE Singapore! platform framework.

- It is difficult to formulate data sharing partnerships such as LIVE Singapore! Our experience in Singapore has been very encouraging because we found highly motivated and cooperative partners but through the program, we have also enhanced our understanding of the necessity to formulate a clear incentive framework for an urban real-time data platform. Simply put: what is the incentive for any one data provider to provide its data to such a platform? And how is it possible to ignite the process when the full potential is unleashed only when there are numerous such data contributors? Again, the experience in Singapore suggests that there is an increased awareness among institutions and companies that the combination of their data with other data sources holds compelling potential, and sharing their data in order to have access to data shared by others is starting to be perceived as a viable approach. Some current data-sharing partners of LIVE Singapore! have expressed interest in sharing some of their data publicly; they have spoken of setting aside certain other data sets on a "reserved" part of the platform. These suggestions have led us to consider modifying the design of the platform's structure.

- Singapore is a lively business environment and we have seen an eagerness to take a project such as LIVE Singapore quickly towards a city-wide implementation level and explore new business opportunities that this would open up. While such an implementation will require further research and development, it does anticipate the mandatory question of a viable and sustainable business model for a platform such as LIVE Singapore! Will data be made available free of charge by the providers? Shall resulting applications be without charge too? If developers charge for applications that make use of several real-time feeds, will the data providers participate in the earnings? And finally what kind of entity (private, public…) would best be indicated to operate the platform which at present is managed within our academic research group?

A key aspect of the next steps of LIVE Singapore! will be "accessibility." After the initial phase of laying the basis and developing a working prototype of the platform architecture, activating real-world, real-time feeds together with our partners as well as programming first applications on top of this framework, LIVE Singapore!'s explicit emphasis is on creating easy access for external data providers as well as easy access to developers and users of the city's real-time streams and allowing for the programmability of these. A second main aspect will be the involvement of other forms of data contribution to the platform such as user generated data. This will bring up the issues of data reliability when moving outside acknowledged institutionalized reliability measures. It will also bring up the opportunity of such a platform in allowing the city's inhabitants to determine new kinds of data to be created that go beyond the existing bits that are generated by existing infrastructures, in this way representing an active and direct involvement of citizens in exploring the future of programming the city.

Note

1. SENSEable City Lab in 2009 organized the "Engaging Data Forum" with the objective of exploring the novel applications for electronic data and addressing the risks, concerns, and consumer opinions associated with the use of these data.

Note on Contributors

Kristian Kloeckl is a research scientist at the Senseable City Lab at the Massachusetts Institute of Technology and is the project leader of the LIVE Singapore! project.

Oliver Senn is a senior research engineer at the Singapore-MIT Alliance for Technology.

Carlo Ratti is director of the Senseable City Lab at the Massachusetts Institute of Technology and is the principal investigator of the LIVE Singapore! project.

Bibliography

F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, "Real-Time Urban Monitoring Using Cell Phones: A Case-Study in Rome," *IEEE Transactions on Intelligent Transportation Systems* 12:1 (2011) 141–151.

F. Calabrese and C. Ratti, "Real Time Rome," Networks and Communication Studies: Official Journal of the IGU's Geography of Information Society Commission 20:3-4 (2006) 247–258.

F. Calabrese, K. Kloeckl, and C. Ratti, "Wikicity: Real-Time Location-Sensitive Tools for the City," paper presented at *Computers in Urban Planning and Urban Management* (Iguassu Falls, July 11–13, 2007).

Changi Airport Group, *About Changi Airport* (2012) <http://www.changiairport.com/our-business/about-changi-airport;jsessionid=C922B544B9134D218744AC97A78131D4>

data.gov <http://www.data.gov/about> Accessed January 3, 2011.

M. Foth, N. Odendaal, and G. Hearn, "The View from Everywhere: Towards an Epistemology for Urbanites," paper presented at the 4th International Conference on Intellectual Capital, Knowledge Management and Organizational Learning (Cape Town, South Africa, 2007).

J. Forrester, *Industrial Dynamics* (Cambridge, MA: MIT Press, 1961).

C. Henson, A. Sheth, and S. Sahoo, "Semantic Sensor Web," *IEEE Internet Computing* 12:4 (2008) 78–83.

IDA Singapore, *Telecommunications Report* (2012) <http://www.ida.gov.sg/Publications/20070822130650.aspx> Accessed 30 March, 2012.

ITU, *The Internet of Things*, Report of the International Telecommunication Union (Geneva, 2005).

V. Mayer-Schönberger, *The Virtue of Forgetting in the Digital Age* (Princeton: Princeton University Press, 2011).

D. Meadows, *Thinking in Systems: A Primer* (White River Junction, VT: Chelsea Green, 2008).

R. Murty, G. Mainland, I. Rose, A. Chowdhury, A. Gosain, J. Bers, and M. Welsh, "CitySense: A Vision for an Urban-Scale Wireless Networking Testbed," paper presented at IEEE International Conference on Technologies for Homeland Security (Waltham, MA, May 12–13, 2008).

Open Geospatial Consortium <http://www.opengeospatial.org> Accessed March 30, 2012. *Oxford English Dictionary* (Oxford: Oxford University Press, 2007).

Pachube <http://www,pachube,com/> Accessed March 1, 2011.

Paulsen, H. and Riegger, U. (2006) "SensorGIS - Geodaten in Echtzeit," In: *GIS-Business* (April 2006).

F. Rojas, K. Kloeckl, and C. Ratti, "Dynamic City, Investigations into the Sensing, Analysis and Application of Real-time, Location-based Data," paper presented at the First International Conference on Critical Digital: What Matters(s)?, (Cambridge, MA, April 18–19, 2008).

F. Rojas, C. Valeri, K. Kloeckl, and C. Ratti, *New York Talk Exchange* (Cambridge, MA: SA+P Press, 2008).

Singapore Department of Statistics, *Time Series on Annual GDP at Current Market Prices* (2012) <http://www.singstat.gov.sg/stats/themes/economy/hist/gdp2.html Accessed July 2012.

Singapore Land Transport, *Statistics in Brief* (2012) <http://www.lta.gov.sg/content/dam/lta/Corporate/doc/LTA-Stats%20in%20Brief%202012_C.pdf> Accessed July 2012.

M. Taylor, *The Moment of Complexity: Emerging Network Culture* (Chicago: University of Chicago Press, 2001).

W3C, *OWL: Web Ontology Language* (2004) <http://www.w3.org/TR/owl-features/> Accessed March 30, 2012.

W3C, *WebSocket API* <http://dev.w3.org/html5/websockets/> Accessed March 30, 2012.

D. Weitzner, H. Abelson, T. Berners-Lee, J. Feigenbaum, J. Hendler, and G. Sussman, "Information Accountability," *Communications of the ACM* 51:6 (2008) 82–87.

Yahoo Pipes <http://pipes,yahoo,com/pipes/> Accessed March 1, 2011.

M. Zardini, ed., *Sense of the City: An Alternate Approach to Urbanism* (Baden, Switzerland: Lars Müller, 2006).

ZeroMQ <http://www.zeromq.org/> Accessed March 3, 2011.