Smart cities will need big data

A new academic center in New York City joins other efforts to harness data from disparate sources with the goal of improving city life.

Ever try catching a cab in Manhattan when it's raining? It turns out that which direction you turn when leaving a building can improve your chances by an order of magnitude. A smartphone application, CabSense, can help you figure that out. It's one small example of urban informatics, a multidisciplinary field that has formed at the intersection of big data and big cities.

A new center at New York University (NYU) will be among the largest efforts thus far to bring big data to bear on learning about and improving the dynamics of big cities. The university's Center for Urban Science and Progress (CUSP) aims to take the pulse of New York City through a wide range of sensors and sift through the resulting torrent of information to improve life in Gotham.

Other centers of urban informatics have cropped up in the US and abroad. For example, two programs are under way at MIT, and the University of Chicago is collaborating with the host city on urban renewal and neighborhood stabilization. The port city of Santander, Spain, has been instrumented with thousands of stationary and mobile sensors, and the Live Singapore! program is providing residents of that city-state with real-time transportation visualizations.

Still, Steven Koonin, the physicist who is director of CUSP, says the center stands alone in terms of scale. At full strength, it will have a budget of $70 million from a combination of federal, state, and city agency support; industry funding; philanthropic donations; tuition; and NYU investments. From an inaugural class of 25 master's students this fall, CUSP is committed to grow to 430 master's and 100 PhD students over a decade.

"What distinguishes us is the fusing of the physical, biological, and informational sensing of the city. In part we're a bunch of physicists, and a lot of these other centers don't have people of that background and orientation," Koonin says. CUSP has 10 industry partners, including Microsoft, Xerox, and IBM; participation from four national laboratories; and partnerships with five academic institutions. Thirteen city agencies plus the Port Authority and the Metropolitan Transportation Authority are cooperating and are holding discussions over the types of data they will provide. The center will "clean up" publicly available government records and make them interoperable.

In addition to fixed *in situ* sensors that record light, temperature, pollution, and other environmental factors, the CUSP team anticipates data sources such as personal sensors like Fitbit and Up wristbands that record the location, activity, and physiology of individual volunteers. Potential internet data sources include Twitter feeds, blogs, other social media, and news articles. Crowd-sourced sensing of the environment and infrastructure could be provided by mobile phones. And a proliferation of video cameras and RFID (radio frequency identification) technology are already monitoring the movement of pedestrians and vehicles. Essentially none of that surveillance data are analyzed for purposes other than forensics or revenue, Koonin says. Plans at CUSP are to fund the installa-

tion and maintenance of sensors that are commercially available but not yet in place.

Potential benefits of urban informatics include real-time systems monitoring, management, and optimization. Results could include managing the flow of traffic, gas, water, and electricity; monitoring the condition of bridges and pipes; planning new public transport routes and utility distribution systems; monitoring public health; and managing emergency response, says Koonin.

The potential contribution of physicists to urban informatics is multifold, says Koonin, a former undersecretary of energy. Physicists are familiar with many of the potential sensor types that would be valuable in the urban setting, notably imaging—"optical, hyperspectral, name your wavelength," he says. Physicists also are proficient at modeling, which, for a city, "is not unlike the challenge of building a complex general climate model," he says. "You've got transportation, you've got economics, communications, health and nutrition, and so on. The challenge is to combine all of those things together to build a good predictive model."

"Physics is an attitude as well as a subject," Koonin continues. "The kind

Steven Koonin, director of New York University's Center for Urban Science and Progress, at a ribbon-cutting ceremony at the center's Brooklyn offices in April. Looking on are NYU president John Sexton (left) and New York mayor Michael Bloomberg.

SAM HOLLENSHEAD FOR NYU

of skills physicists bring to thinking through complicated situations, data driven and so on, are not all that common in urban science and technology at this point. Physicists have a lot to bring to the table here."

Parking and streetlights

In the urban informatics arena, CUSP is only the most recent entry. The port of Santander on the northern coast of Spain is the most data-intensive city in Europe. With a grant of €8 million ($10.6 million) from the European Commission three years ago, a consortium of European universities and telecom companies has installed some 18 000 stationary and mobile sensors of various types throughout the municipality of around 180 000 residents. In addition to monitoring air pollution, noise, and other environmental conditions, sensors indicate when dumpsters require emptying and when streetlights have burned out or can be switched off because nobody is around. Sensors buried in the pavement detect open parking spaces and relay that information to digital displays mounted at major intersections to help guide drivers.

University of Cantabria engineering professor Luis Muñoz, principal investigator for the SmartSantander project, says the goal from the outset has been to make the sensing infrastructure available to researchers while simultaneously providing useful services to the city. SmartSantander also features a smartphone application that allows residents to report problems such as potholes and to track the city's response. Using their smartphones, residents can use an "augmented reality" system comprising 2600 optical and wireless tags at tourist attractions, shops, bus stops, and other locations throughout the city to readily obtain online information about those locations.

The Urban Center for Computation and Data in Chicago is working with a large collection of data supplied by the city. "We're taking that data and defining two different kinds of people to look at it," says center director Charlie Catlett. "One are the computational scientists and data analytics people," he says, and the other group comprises "the social, behavioral, and economic scientists, and particularly those that are used to dealing with data, but not so much data."

Some of the data, such as the GPS tracks of public transit vehicles, are public. The city also makes available to the center data such as education and incarceration records with personal identifiers

stripped, data from gunshot sensors, and the GPS tracks of police vehicles.

A second thrust of the center, Catlett says, is evidence- and science-based planning and design of cities. That's largely motivated by the 600-acre Lakeside redevelopment project now in the design stage at the site of a demolished steel mill in Chicago. The project architects, who are partners in the center, use software tools to determine the need for electricity and water, transportation, stormwater management, and other infrastructure. "What we want to do is take the data out of that design and couple it with computational models" of energy and water demand in buildings and of transportation into and out of the development, Catlett says.

Ultimately, Catlett and his colleagues want to do what he calls "predictive analytics," using big data to help alert city officials to neighborhoods that are just beginning to decline, so that actions can be taken to reverse the trend. Indicators such as emergency calls, crime reports, building permits and inspections, health reports, and tax receipts could point to neighborhoods on the edge. The Cook County Land Bank, which buys up abandoned properties to reduce the negative impact on the surrounding area, might use such predictive capability to help make the most of its scarce resources, he explains.

Formula One

"What is happening at an urban scale today is similar to what happened two decades ago in Formula One auto racing," says Carlo Ratti, who heads MIT's Senseable City Laboratory. "Up to that point, success on the circuit was primarily credited to a car's mechanics and the driver's capabilities. But then telemetry technology blossomed. The car was transformed into a computer that was monitored in real time by thousands of sensors, becoming intelligent and better able to respond to the conditions of the race. In a similar way, over the past decade, digital technologies have begun to blanket our cities, forming the backbone of a large, intelligent infrastructure."

Broadband fiber-optic and wireless telecommunications grids are supporting increasingly affordable mobile phones and tablets, Ratti notes. Meanwhile, open databases—such as the federal government's Data.gov site—are revealing all kinds of information. "Add to this foundation a relentlessly growing network of sensors and digital-control technologies, all tied together by cheap, powerful computers, and our

cities are quickly becoming like computers in open air," he says.

The Senseable City Lab also operates the Live Singapore! project, which is developing an open platform for the collection, synthesis, and distribution of real-time data that originate from a large number of different sources. The goal is for developers to build multiple applications that will extract usable information from the data. While the platform is still being built, Ratti says a startup company is now working on an open application using real-time flight data to better manage taxi traffic.

Alex Pentland directs the Human Dynamics Laboratory, also at MIT. He notes that cities have been built to accommodate peak infrastructure demands, but little is known about actual demand or its patterns. With electricity, for example, "we just deliver by pouring electrons into the thick wire, and then hope it all goes out." Where the power is actually used isn't known. "Imagine how much better you could make a system if you could figure out where it was being used."

For power, supplying the last 5% of peak demand is far more expensive and dirtier than supplying the first 95%, since older, less efficient plants are usually brought on line last. "If you could anticipate demand, you could shut down some parts and raise up other parts [of the system]. You could save an enormous amount of energy and the pollution that goes with it. But you have to know where people are and where they are going next to be able to do that," Pentland says.

Similarly, if one can track the actual numbers of people flowing from, say, New York City's Penn Station to the lower east side of Manhattan and see how that pattern differs from day to day, then transit systems could be reconfigured accordingly to make them more efficient, he says.

Cell phones are ubiquitous and can be a rich source of data on the dynamics of cities. Pentland notes that "if there's a public need for [that data], it's trivial to get it. But the public has to see the need, and there has to be security in the sense that it won't be abused and it really is anonymous."

Pentland helped develop the CabSense application, which is owned by the MIT–Columbia University spinoff Sense Networks. CabSense uses driver logs from more than 200 million taxi pickups to predict the best corners to hail a cab. The ratings change based on time of day and day of the week.

David Kramer